Science Library

Vol. 36, No. 2

MAY 8 1946

April, 1946

Journal of Experimental Psychology

EDITED BY

SAMUEL W. FERNBERGER, ON LEAVE
FRANCIS W. IRWIN, ACTING EDITOR
UNIVERSITY OF PENNSYLVANIA

CONSULTING EDITORS

CHARLES W. BRAY, *Princeton University;* ELMER K. CULLER, *University of Rochester;* CLARENCE H. GRAHAM, *Brown University;* JOY P. GUILFORD, *University of Southern California;* FRANCIS W. IRWIN, *University of Pennsylvania;* DONALD G. MARQUIS, *Yale University;* ARTHUR W. MELTON, *University of Missouri;* CARROLL C. PRATT, *Rutgers University.*

CONTENTS

Eye-Movements during Prolonged Reading: A. C. HOFFMAN 95

The Constancy of Colored Objects in Colored Illumination: H. WALLACH AND ALICE GALLOWAY 119

The Width of the Basilar Membrane in the Guinea Pig: S. PERLIN 127

The Relation between Frequency of Blinking and Effort Expended in Mental Work: M. E. BITTERMAN AND E. SOLOWAY 134

The Effect of Negative Incentives in Serial Learning. I. The Spread of Variability under Electric Shock: G. R. STONE 137

Genetic Changes in Semantic Conditioning: B. F. RIESS 143

Incidental Learning and Generality of Set: L. POSTMAN AND VIRGINIA L. SENDERS 153

The Learning and Retention of Concepts. II. The Influence of Length of Series. III. The Origin of Concepts: H. B. REED 166

Discussion: Comments on Criteria of Ease of Reading: M. LUCKIESH, 180. *A Reply to Dr. Luckiesh:* M. E. BITTERMAN, 182. *Reynolds' Repetition of Blodgett's Experiment on Latent Learning:* H. C. BLODGETT, 184.

PUBLISHED BI-MONTHLY

BY THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

PRINCE AND LEMON STS., LANCASTER, PA.

AND 1227 NINETEENTH STREET, N. W., WASHINGTON 6, D. C.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879

PUBLICATIONS OF

THE AMERICAN PSYCHOLOGICAL ASSOCIATION

WILLARD L. VALENTINE, *Business Manager*

PSYCHOLOGICAL REVIEW

HERBERT S. LANGFELD, *Editor*
Princeton University

Contains original contributions only, appears bi-monthly, January, March, May, July, September, and November, the six numbers comprising a volume of about 540 pages.
Subscription: $5.50 (Foreign, $5.75). Single copies, $1.00.

PSYCHOLOGICAL BULLETIN

JOHN E. ANDERSON, *Editor*
University of Minnesota

Contains critical reviews of books and articles, psychological news and notes, university notices, and announcements. Appears bi-monthly beginning in January. Special issues of the BULLETIN consist of general reviews of recent work in some department of psychology.
Subscription: $7.00 (Foreign, $7.25). Single copies, $1.25.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

S. W. FERNBERGER, *Editor, on Leave,* FRANCIS W. IRWIN, *Acting Editor*
University of Pennsylvania

Contains original contributions of an experimental character. Appears bi-monthly (beginning 1944), one volume per year, each volume of six numbers containing about 520 pages.
Subscription: $7.00 (Foreign, $7.25). Single copies, $1.25.

PSYCHOLOGICAL ABSTRACTS

WALTER S. HUNTER, *Editor*
Brown University

Appears monthly, the twelve numbers and an index supplement making a volume of about 700 pages. The journal is devoted to the publication of non-critical abstracts of the world's literature in psychology and closely related subjects.
Subscription: $7.00 (Foreign, $7.25). Single copies, 75c.

PSYCHOLOGICAL MONOGRAPHS

JOHN F. DASHIELL, *Editor*
University of North Carolina

Consist of longer researches or treatises or collections of laboratory studies which it is important to publish promptly and as units. The price of single numbers varies according to their size. The MONOGRAPHS appear at irregular intervals and are gathered into volumes of about 350 pages.
Subscription: $6.00 per volume (Foreign, $6.30).

JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY

GORDON W. ALLPORT, *Editor*
Harvard University

Appears quarterly, January, April, July, October, the four numbers comprising a volume of 400 pages. The journal contains original contributions in the field of abnormal and social psychology, reviews, notes and news.
Subscription: $5.00 (Foreign, $5.25). Single copies, $1.50.

JOURNAL OF APPLIED PSYCHOLOGY

DONALD G. PATERSON, *Editor*
University of Minnesota

Covers the applications of psychology in business, industry, education, etc. Appears bi-monthly, February, April, June, August, October, and December.
Subscription: $6.00. Single copies, $1.25.

Subscriptions, orders, and business communications should be sent to

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

1227 NINETEENTH STREET, N. W., WASHINGTON 6, D. C.

Journal of
Experimental Psychology

VOL. 36, NO. 2 APRIL, 1946

EYE-MOVEMENTS DURING PROLONGED READING *

BY ARTHUR C. HOFFMAN

Tufts College

It has been observed in clinical situations that patients complain of being unable to continue reading for extended periods of time. No single investigation can hope to give an explanation of this seemingly simple statement on the part of the patient. The complex nature of both cause and expression of this complaint precludes a conclusive answer based on a single experiment. It is possible, however, to subject individuals to prolonged reading under experimental conditions and to observe a portion of their behavior. Modern electrical instruments enable accurate recording of the movements of the eyes while reading—movements such as fixations, regressive movements, and blinking. The experimenter could thus observe the changes in the functioning of the oculomotor system as they may occur during continued activity. It is recognized that reading is a complex of psychological and physiological functions. The experiment to be described, however, concerns itself only with certain observed patterned movements of the extrinsic oculomotor system occurring under conditions of prolonged reading.

A review of the scientific literature on receptor activity reveals that experimentation has been concerned primarily with phenomena following changes in stimulus characteristics or variations introduced in the sensory mechanism itself. Seldom has the experimental problem been concerned with changes in receptor function when time alone has been the variable, in the sense of prolonging a function of the receptor organ and noting the progressive effects of such continued activity.

The above statement has primary reference to the literature on ocular functions involved in reading. Experimenters have usually taken great care, and rightly so, to eliminate, or at least to mitigate,

* This paper is an abridgment of a thesis submitted to the University of Rochester in partial fulfillment of the requirements for the Ph.D. degree.

wherever possible the interference of temporal variables—concern with appropriate adaptation, practice effect, and length of experimental sitting are cases in point. But, in addition to academic interest in effects which might occur during prolonged receptor activity, practical considerations require scientific data to understand and cope with the problems arising from the continued use of the eyes. The amount of visual work imposed by modern educational methods and the format of the present social milieu have tended to increase the importance of the eyes and have made even more necessary an investigation of the condition commonly referred to as 'visual or ocular fatigue.'

If it is demonstrated that there are significant changes or decrements in ocular functions as a result of continued activity, the problem remains as to the nature of this reality. Where such decrements are found in behavioral functions, they have been imputed to 'fatigue.' Contemporary replies to the question of what is visual fatigue or what is its nature have been, in the author's opinion, primarily assumptions and hypotheses, since sufficient scientific data are lacking.

In the present experiment no explicit assumptions are made at the outset as to the nature of the changes which may be demonstrated, either as to their causes or their characteristics. This experimental attitude does not, of course, preclude the privilege of drawing inferences from the data gathered in this experiment. The primary purpose of the investigation, however, is simply to gather data from which explanatory definitions can eventually be made. Nevertheless, by implicit assumption and by usual definition—the fact that the experimental variable is continued activity from which a work decrement might result—the experimental problem investigated finds its orientation in the general field of fatigue studies.

Within the knowledge of the present author, experiments involving both time and reading as variable have had what in the present instance would be the ulterior purpose of investigating other variables as well. In anoxemia experiments by McFarland, Knehr, & Berens (**10**) 10 *Ss* were tested in an oxygen chamber for an hour at a time under atmospheric conditions of outside air, of 12.5 percent oxygen, and of 10.5 percent oxygen. In each experimental situation the *Ss* read six paragraphs, during which time eye-movements were photographically recorded. The first two paragraphs were read 15 min. after entering the chamber, the second pair after one-half hour, and the third at the end of the hour. The measures of ocular performance studied were reading time, number of fixations, number of regressions per line, and binocular adjustments at the beginning and end of reading a line of print and during fixation pauses. These measures

during each of the three experimental situations were compared. Reading time and number of fixations in the 10.5 percent oxygen situation were significantly greater than under the normal condition. The average binocular adjustment during fixations decreased significantly. Variability among *Ss* increased only slightly; *Ss* also tended to maintain their same relative position from one experimental condition to the next. Qualitative changes in the eye-movement record suggested that the precision of movement and the maintenance of fixations decreased in efficiency. During an experiment by Clark and Warren (3) three college students maintained a sixty-five-hour vigil. Among other tasks, the *Ss* read short selections every 10 hours. The eye-movement records showed no uniform change in the number of fixations and regressions per line, binocular adjustments, or reading time, although, in some instances performance scores were higher on the final tests. Eye-movement records of an *S* were made by Dearborn (4) at the beginning and end of a day's proof reading. No evident change in number of fixations was noted, but reading time was found to increase.

These several experiments suggest indirectly that changes in oculomotor efficiency might be expected to occur with prolonged activity.

Experimental Procedure

In the opinion of the author a direct approach to the problem of prolonged reading requires an experimental procedure in which (1) a sufficient number of *Ss* (2) read continuously (3) for a prolonged period of time. The procedures of the above experiments involve one or more, but not all, of these requisites. In the present experiment to be reported 30 *Ss* read continuously for four hours, during which time continuous records of eye-movements were made.

The ocular functions observed in this experiment involve the activity of the six extrinsic muscles of the eyeball and the muscles of the eyelid, as well as the nervous mechanisms activating these muscle groups. The particular ocular movements investigated were those of the reading pattern: the blinking movements of the eye, fixation pauses, the lines of print read, and the regressive movements of the eye.

A blink may be defined as a temporary closure of both eyes, involving movements of the upper and lower eyelids, so that the pupil is momentarily hidden from view. However, Miles (11) has demonstrated that concomitant with the closure of the lids, the eyeball turns upward. It is this upward movement of the eye during a blink that has been recorded in the present experiment.

The term fixation has the general meaning to look at a stimulus object without changing the lines of regard; as an ocular function measured in this experiment, the term refers to the specific case of fixation as a component movement of the eye-movement reading pattern. A fixation (or frequently, a fixation pause) may be described as the time in reading a line of print during which the eyeball is held stationary and the retina is exposed to the stimulus pattern.

Normally, reading proceeds from left to right along the line of print—a fixation pause followed by a short saccadic movement of the eyeballs to the right, another fixation pause, another saccade to the right, etc. There may occur, however, regressive (or corrective) movements, which as the term suggests are saccadic movements of the eyeball to the left, allowing re-fixation of material earlier in the line of print.

These ocular functions were observed in 30 Harvard undergraduates ranging in age from 18 to 26 years, with a median age of 20 and a mean age of 20.2. They were paid $2.50 at the end of the trial period. Nine of the *Ss* wore glasses always, eight wore glasses only while reading, and 13 did not wear glasses. The *Ss* read steadily for four hours *The story of civilization* by George A.

Dorsey. The edition used was printed on creme-colored paper in 11-point old style linotype with 26 picas to the line and one point of leading. The length of the printed page was 42 picas over-all.

Because of the sensitivity of the recording apparatus used, it was necessary for the *Ss* to read in an electrically shielded room. This specially built chamber was four feet wide, eight feet

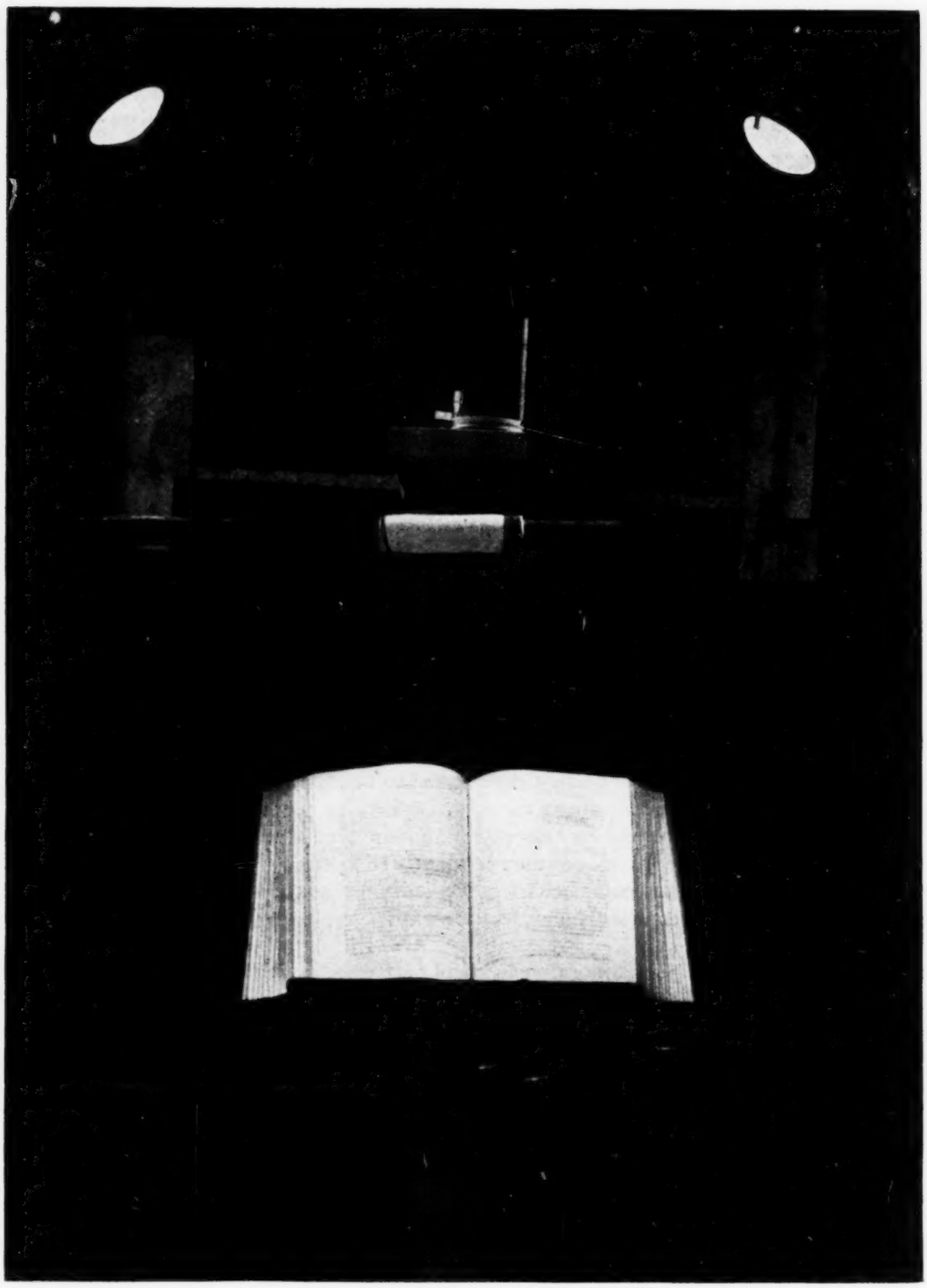

FIG. 1. Front view of the shielded reading room

high, and eight feet long. Fig. 1 shows a view from the door of the chamber; a side view from the window to the *Ss* left is shown in Fig. 2. (The woodwork seen in the figures was for ocular tests not described in this paper.) The walls were painted black. The *S* read at a table (also covered

in black) on which was placed a book-stand with a large black paper-board behind it. Once the book-stand had been adjusted to the tilt and distance preferred by the *S*, it was not changed during the four-hour reading period. Disturbance from the laboratory was kept at a minimum, so that the atmosphere of the experimental chamber was that of a reading cubicle in a university

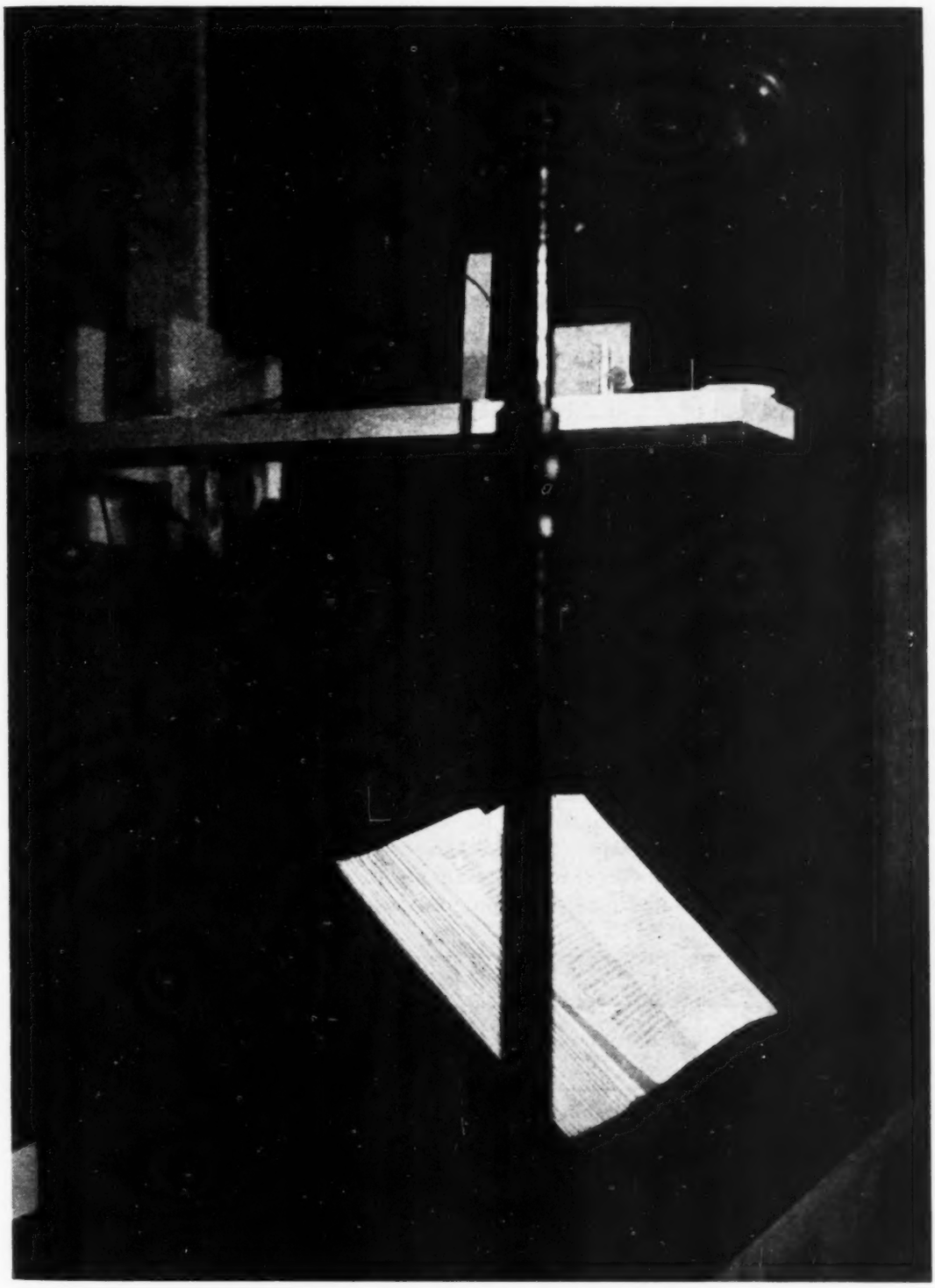

FIG. 2. Side view of the shielded reading room

library. The illumination on the book of 10.5 footcandles came from three light sources above the *S*.

Each experimental trial began at two o'clock in the afternoon; the four-hour reading period began around 2:30 P.M. The *S* read a preliminary book in the experimental chamber for from

10 to 15 min. before the experiment proper began, so as to accustom himself to the experimental situation and to allow time for adjusting the apparatus. The *S* was instructed to read steadily without interruption and to disregard all movements or noises in the outside laboratory.

The behavior of the ocular functions during the whole four-hour period was indirectly observed by electrically recording the changes in the corneo-retinal potential as the eyeball was moved in its socket—the electroculogram. The reliability and validity of this method of recording eye-movements has been shown to be adequate in previous experiments by Hoffman, Weilman, & Carmichael (7), Mowrer, Ruch, & Miller (14), Miles (12, 13), Lindsley & Hunter (9), and others.

The apparatus used included electrodes, electrode board, pre-amplifier, power amplifier, and oscillograph units. The changes in potential were picked up by silver-hat electrodes attached to the skin about the eyes of the *S*. Four electrodes were used in this experiment: one beside each external canthus, one over the right eyebrow, and one just below the right eye. The first pair picks up the potential changes due to the horizontal movements of the eyeballs, the second pair reacts to the vertical movements, e.g., blinking. The lead wires from these electrodes were plugged into an electrode board behind the *S*, which was the only part of the recording apparatus in the *S*'s immediate environment. The wires were long enough to permit unrestrained movement on the part of the *S*.

Control switches of the pre-amplifier made it possible to introduce to the channels of the amplifier any pair of electrodes desired. Another pre-amplifier control switch made it possible to determine the amount of potential causing so much excursion on the record. The amplifiers used were of the paired non-interfering, resistance-capacity coupled type. This equipment is capable of recording the absolute magnitude of potential shifts, but not the recording of maintained potentials. The amplification of the corneo-retinal potentials could be altered by gain controls on the power amplifier. They cut down the total amplification to a point in the circuit where recording can be accomplished without distortion. The higher the amplification, the smaller, of course, the potential necessary to give a deflection of the ink-writer. There were also two filter switches on the power amplifier; one of which progressively cut out low frequencies, and the other, high frequencies. These switches can be used to reduce the amount of muscle potentials or sixty-cycle interference in the record.

The ink-writer consisted of two individual oscillograph units and the paper-pulling mechanism. The oscillograph units were built on the moving coil principle. The peak-to-peak linear amplitude of each was about one and one-half cm. The units were so arranged that the pens touched the paper only hard enough to produce a smooth, clear-flowing ink-line. The paper-pulling mechanism was so adjusted that the paper moved at a constant rate of one and one-half cm. per sec. The electroculogram made with this entire apparatus is very similar to the records made with the more usual photographic methods, except for the fact that the pens tend to return toward the baseline after making a deflection because of the nature of the apparatus used. The electroculogram, therefore, is less a step-wise progression than a series of peaks, as may be seen in Fig. 3.

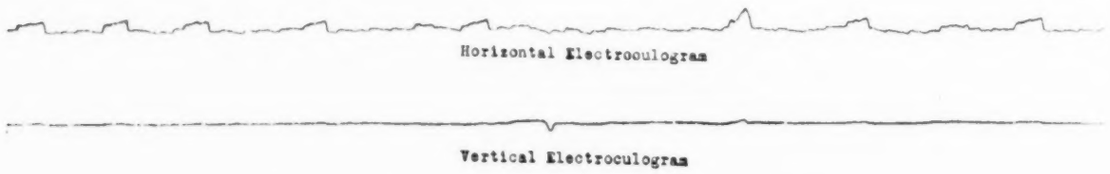

FIG. 3. A sample page from the electroculogram

Nine five-min. samples were taken of each continuous four-hour record; thus, 45 min. of the record were read by the *E*. The sample record included the first five-min. portion of the four-hour period and, from then on, the last five-min. portion of each half-hour. These sample periods will be referred to by the letter 'i' for the initial five min. and by Roman numerals for the five min. at the end of each half-hour: i, I, II, III, IV, . . . VIII. It is to be noted that in taking these samples, neither the *E* nor the reader knew beforehand or during the four-hour reading period which actual part of the continuous record was to be sampled. The sample record was read or interpreted by direct observation, since for the ocular functions investigated, only counting rather than measurement of the record was necessary.

From the electroculogram of each five-min. sample period were determined the number of blinking movements made, the number of fixations, the number of lines read, and the number of regressive movements made. From these data additional data were derived: the number of fixations per line was determined by dividing the number of fixations made during a sample period by the number of lines read; the number of regressions per line by dividing the number of regressions by the number of lines read; and the standard deviation of the number of fixations per line by using the number of fixations per line as the mean and the number of lines read as N. Therefore, for each of the nine sample periods of the records of each of the 30 Ss there were available for analysis seven measures of the reading eye-movement pattern, to be referred to hereafter as: Blinks Number, Fixations Number, Fixations per Line, Fixations Sigma Score, Lines Read Number, Regressions Number, and Regressions per Line.

Results

Should the reader be interested in the actual numerical values of the eye-movements measured in this experiment, graphs of the sample means for five of the seven ocular measures are presented in Figs. 4 and 5. Fig. 4 plots the sample means of those measures which

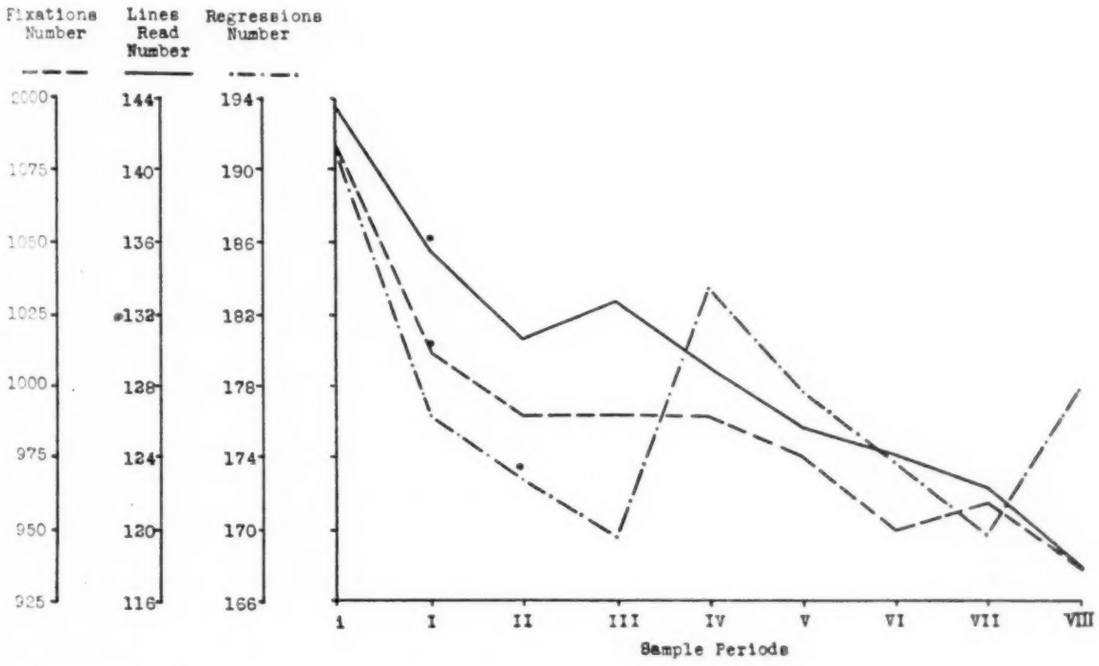

Fig. 4. Graphs of sample means of the three eye-movement measures showing decrement

decreased during the prolonged activity; Fig. 5 shows the sample means of the measures which increased. It will be noted that the numerical width of the unit indicated on the y-axis varies widely from measure to measure. The asterisks on the graphs are used to indicate the earliest sample at which performance was found to be significantly different from that of the initial sample period as indicated by the statistical analysis to be described below.

A preliminary analysis of regression (not shown) by applying orthogonal polynomials (1) to the data for each of the eye-movement measures indicated in the majority of instances that only the first or

linear regression may be regarded as significant. This shows that if any changes occurred during the four-hour period of reading, they were probably uncomplicated by curvilinear relationship with time and could therefore be regarded as more or less 'steady' decrements or increments, as the case might be. Whether or not, however, a more exhaustive analysis of the same sort would support the above conclusion as to linear regression, it was thought to be of more immediate value to present the results of the analysis of variance technique (5). Since a test (6) of the sample variances in the case of each measure showed they could be considered homogeneous, this technique was used to determine whether significant changes occurred

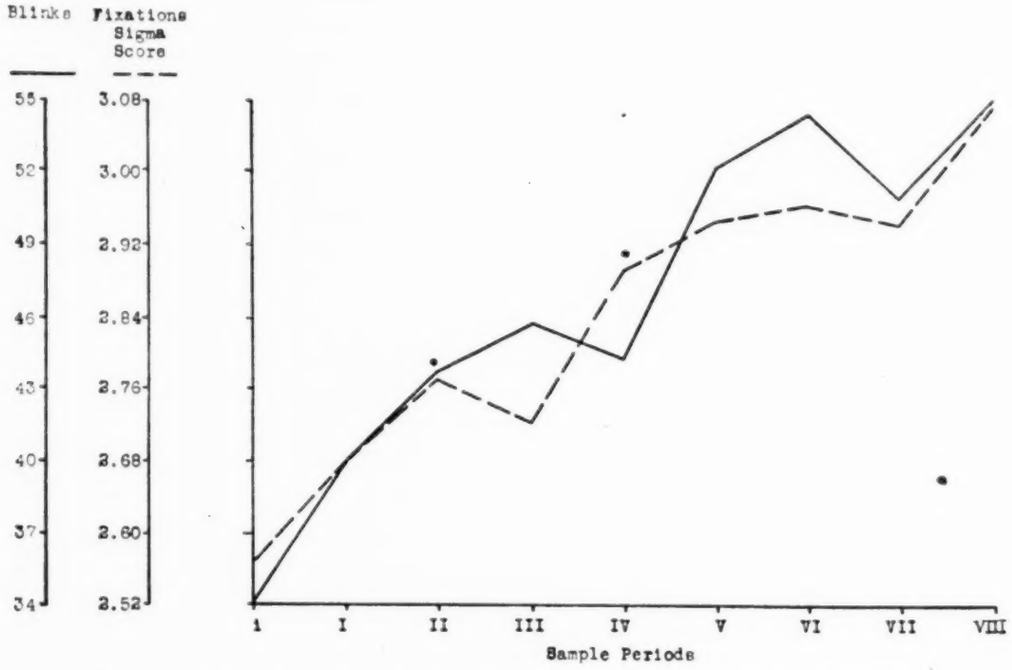

FIG. 5. Graphs of sample means of the two eye-movement measures showing increment

during the four-hour reading period. The results of the analyses (to be described later) of the data on each eye-movement measure are presented in Tables I–VII.

The same numbers of *Ss* (30) and of samples (nine) were involved in securing the data on each of the seven eye-movement measures; the following remarks, therefore, are typical of the statistical procedure used in all seven cases. A total of 269 degrees of freedom were available for each eye-movement measure. Of these, eight degrees of freedom may be assigned to the samples in time, 29 are available for the variance labelled subjects and 232 degrees of freedom remain for the variance ('error' or remainder variance) with which to evaluate the significance of the above variances. It is recognized, of course, that this latter term is the first order interaction, which might be interpreted in the present instance as the variance due to the possible

TABLE I

ANALYSIS OF DATA

Blinking number

	S.S.	n.d.f.	Mean Square
Samples (1)	11,617.0290	8	1,452.1286
Subjects (2)	304,199.5620	29	10,489.6402
Remainder (3)	41,622.3050	232	179.4065
Total	357,438.8960	269	

F-test: $\frac{V_1}{V_3} = 8.094^*$ $\frac{V_2}{V_3} = 58.468^*$ $\frac{V_2}{V_1} = 7.224^*$

t-test

	i	I	II	III	IV	V	VI	VII
I	5.600							
II	9.267*	3.667						
III	11.367*	5.767	2.100			*8.985*		
IV	9.967*	4.367	0.700	−1.400				
V	17.700*	12.100*	8.433	6.333	7.733			
VI	20.000*	14.400*	10.733*	8.633	10.033*	2.300		
VII	16.500*	10.900*	7.233	5.133	6.533	−1.200	−3.500	
VIII	20.633*	15.033*	11.367*	9.267*	10.667*	2.933	0.633	4.133

* Significant at the 1 percent level of confidence.

TABLE II

ANALYSIS OF DATA

Fixations number

	S.S.	n.d.f.	Mean Square
Samples (1)	442,976.2194	8	55,372.0274
Subjects (2)	2,257,783.3750	29	77,854.5991
Remainder (3)	842,482.5584	232	3,631.3903
Total	3,543,242.1528	269	

F-test: $\frac{V_1}{V_3} = 15.248^*$ $\frac{V_2}{V_3} = 21.439^*$ $\frac{V_2}{V_1} = 1.406$

t-test

	i	I	II	III	IV	V	VI	VII
I	−69.567*							
II	−92.867*	−23.300						
III	−92.400*	−22.833	0.467			*40.423*		
IV	−92.667*	−23.033	0.200	−0.267				
V	−106.233*	−36.667	−13.367	−13.833	−13.567			
VI	−134.400*	−64.833*	−41.533*	−42.000*	−41.733*	−28.167		
VII	−124.400*	−54.833*	−31.533	−32.000	−31.733	−18.167	10.000	
VIII	−145.267*	−75.700*	−52.400*	−52.867*	−52.600*	−39.033	−10.867	−20.867

* Significant at the 1 percent level of confidence.

TABLE III

ANALYSIS OF DATA

Fixations per line

	S.S.	n.d.f.	Mean Square
Samples (1)	7.2062	8	0.9008
Subjects (2)	520.3750	29	17.9440
Remainder (3)	54.8116	232	0.2362
Total	582.3928	269	

F-test: $\frac{V_1}{V_3} = 3.814^*$ $\frac{V_2}{V_3} = 75.970^*$ $\frac{V_2}{V_1} = 19.920^*$

t-test

	i	I	II	III	IV	V	VI	VII
I	−0.053							
II	−0.020	0.033						
III	−0.057	−0.003	−0.037			*0.3255*		
IV	0.100	0.153	0.120	0.157				
V	0.230	0.283	0.250	0.287	0.130			
VI	0.137	0.190	0.157	0.193	0.037	−0.093		
VII	0.300	0.353*	0.320	0.357*	0.200	0.070	0.163	
VIII	0.430*	0.483*	0.450*	0.487*	0.330*	0.200	0.293	0.130

* Significant at the 1 percent level of confidence.

TABLE IV

ANALYSIS OF DATA

Fixations sigma score

	S.S.	n.d.f.	Mean Square
Samples (1)	6.1138	8	0.7642
Subjects (2)	76.5304	29	2.6390
Remainder (3)	30.3373	232	0.1308
Total	112.9815	269	

F-test: $\frac{V_1}{V_3} = 5.842^*$ $\frac{V_2}{V_3} = 20.176^*$ $\frac{V_2}{V_1} = 3.453$

t-test

	i	I	II	III	IV	V	VI	VII
I	0.113							
II	0.200	0.087						
III	0.147	0.033	−0.053			*0.2424*		
IV	0.317*	0.203	0.117	0.170				
V	0.367*	0.253*	0.167	0.220	0.050			
VI	0.387*	0.273*	0.187	0.240	0.070	0.020		
VII	0.373*	0.260*	0.173	0.227	0.057	0.007	−0.013	
VIII	0.497*	0.383*	0.297*	0.350*	0.180	0.130	0.110	0.123

* Significant at the 1 percent level of confidence.

TABLE V

ANALYSIS OF DATA

Lines read number

	S.S.	n.d.f.	Mean Square
Samples (1)	14,464.7370	8	1,808.0921
Subjects (2)	175,883.9470	29	6,064.9638
Remainder (3)	30,692.1530	232	132.2938
Total	221,040.8370	269	

F-test: $\frac{V_1}{V_3} = 13.667^*$ $\frac{V_2}{V_3} = 45.845^*$ $\frac{V_2}{V_1} = 3.354$

t-test

	i	I	II	III	IV	V	VI	VII
I	−8.000*							
II	−12.767*	−4.767						
III	−10.800*	−2.800	1.967			*7.716*		
IV	−14.500*	−6.500	−1.733	−3.700				
V	−17.967*	−9.967*	−5.200	−7.167	−3.467			
VI	−19.400*	−11.400*	−6.633	−8.600*	−4.900	−1.433		
VII	−21.400*	−13.400*	−8.633*	−10.600*	−6.900	−3.433	−2.000	
VIII	−25.867*	−17.867*	−13.100*	−15.067*	−11.367*	−7.900*	−6.467	−4.467

* Significant at the 1 percent level of confidence.

TABLE VI

ANALYSIS OF DATA

Regressions number

	S.S.	n.d.f.	Mean Square
Samples (1)	11,161.0046	8	1,395.1256
Subjects (2)	1,042,647.3146	29	35,953.3561
Remainder (3)	116,026.5554	232	500.1145
Total	1,169,834.8746	269	

F-test: $\frac{V_1}{V_3} = 2.790^*$ $\frac{V_2}{V_3} = 71.890^*$ $\frac{V_2}{V_1} = 25.771^*$

t-test

	i	I	II	III	IV	V	VI	VII
I	−14.433							
II	−18.033*	−3.600						
III	−21.567*	−7.133	−3.533			*15.001*		
IV	−7.733	6.700	10.300	13.833				
V	−13.267	1.167	4.767	8.300	−5.533			
VI	−17.100*	−2.667	0.933	4.467	−9.367	−3.833		
VII	−21.233*	−6.800	−3.200	0.333	−13.500	−7.967	−4.133	
VIII	−13.400	1.033	4.633	8.167	−5.667	−0.133	3.700	7.833

* Significant at the 1 percent level of confidence.

TABLE VII

ANALYSIS OF DATA

Regressions per line

	S.S.	n.d.f.	Mean Square
Samples (1)	1.2200	8	0.1525
Subjects (2)	96.1325	29	3.3149
Remainder (3)	9.8977	232	0.0427
Total	107.2502	269	

F-test: $\frac{V_1}{V_3} = 3.571^*$ $\frac{V_2}{V_3} = 77.632^*$ $\frac{V_2}{V_1} = 21.737^*$

t-test

	i	I	II	III	IV	V	VI	VII
I	−0.033							
II	−0.014	0.019						
III	−0.065	−0.032	−0.051			*0.1386*		
IV	0.072	0.105	0.087	0.137				
V	0.054	0.087	0.068	0.119	−0.019			
VI	0.063	0.096	0.077	0.128	−0.010	0.009		
VII	0.048	0.081	0.062	0.113	−0.025	−0.006	−0.015	
VIII	0.175*	0.208*	0.189*	0.240*	0.102	0.121	0.112	0.127

* Significant at the 1 percent level of confidence.

differential effect of prolonged reading upon the individual subjects. It is likely that individuals differ in their susceptibility to 'fatiguing' conditions, and evidence for such a conclusion would be a valuable contribution toward a description of fatigue. Since, however, the time factor is of primary interest in the present research, and the effect of the prolonged reading is presumably measured by the variance assigned to samples, the interaction variance was used as a measure of 'error.' If the sample variance proves to be significantly greater than this 'error' variance, it would not necessarily mean that there was no significant interaction between continued activity and individual subjects, but rather that, by means of the present experimental design, evidence was found for concluding that the variance assigned to the time samples was by itself an important consideration.

In the case of all seven eye-movement measures the remainder variance (V_3) was found to be smaller than the amount of variance assigned to samples (V_1) or to subjects (V_2). Also in every case, the variance due to samples was smaller than the variance assigned to *Ss*. Three variance ratios were therefore computed: sample variance over remainder variance (V_1/V_3), subjects over remainder (V_2/V_3), and subject variance over sample variance (V_2/V_1) The *F*-test was used to test the significance of each of these ratios. The value of *F* appearing by chance once in a hundred times is chosen as the level of

significance (i.e., the one percent level of confidence). The values which the observed ratios must exceed in each case to be regarded as indicative of significant difference at this level of confidence are presented below:

Ratio	1%
V_1/V_3	2.59
V_2/V_3	1.79
V_2/V_1	5.20

In presenting the analyses of the data for each eye-movement measure, variance ratios regarded as significant are starred.

In the case of each measure the variance assigned to samples was significantly greater than the remainder variance, indicating that changes in ocular functioning occurred during the prolonged reading. Since the purpose of this experiment was to observe eye-movement performance in time, it was of further interest to determine with the samples at hand the point during the four-hour period at which changes or differences from initial performance become of such magnitude that they may be regarded as significant. For this purpose the *t*-test (8) was applied to evaluate the differences between the means of the nine individual samples for each eye-movement measure. The remainder term (with 232 degrees of freedom) of the previous analyses was again used as a measure of error.

At the one percent level of confidence, a *t*-ratio under these conditions must exceed 2.598 to indicate a significant difference. It was more convenient in the present experiment, however, to compute for each eye-movement measure the value which an observed difference between two sample means must exceed to be regarded as significant. In Tables I–VII, therefore, are also presented differences between sample means (i, I, II, etc., designate the sample means for which the appropriate difference appearing in the table has been computed). The mean of an earlier sample has been subtracted from the mean of a later sample ($M_l - M_e$); thus a positive difference indicates an increase in the later mean and a negative value, a decrement. The difference value to be exceeded is shown in the upper right-hand corner of each table. Differences in the table which may be regarded as significant are again starred. Primary interest, of course, is in the significant differences that may appear in the first column of each table in which the mean of the initial sample is compared to the means of all subsequent samples.

Blinking number

The analysis of data for the number of blinks measure appears in Table I. The variance due to samples and to *Ss* is significantly greater than the remainder variance. The variance due to *Ss* is significantly greater than the variance assigned to samples. In

evaluating blinking performance, therefore, these ratios show that it is necessary to know both who is doing the reading and how long a particular individual has been reading, with the further indication that the former consideration is the more important. The differences between the mean of the first sample period (i) and the mean of each subsequent sample period increase. The progression is not regular, but the direction of increase is definite. These differences are significant after one hour of reading (sample II). At the time of significant change the mean number of blinks has increased from 34.3 to 43.6, or 27 percent more than the number made during the first sample period. When subsequent means are compared to the mean performance at the end of the first half-hour (I), significant change occurs after two and one-half hours of reading (V).

Fixations number

The analysis of the data for the number of fixations measure is presented in Table II. The first two variance ratios shown are significant, the third is not. These findings show that both the length of time and the individuals doing the reading are equally important factors influencing fixation performance. Individual differences are not a more important consideration than the time spent in reading, but prolonged activity evidently does not eliminate individual differences. The differences between the mean of the initial sample and those of subsequent sample periods show that the number of fixations made decreases in fairly regular order. This decrement is significant after one-half hour of reading (as shown, of course, by the time intervals sampled in this experiment). The mean differences shown begin at a six percent decrement (1012.9–1082.5) and end after four hours of reading at a 13 percent decrement (937.2–1082.5). Comparison of sample means with the mean performance during I shows significant differences after three hours of reading (VI).

Fixations per line

Table III presents the analysis of the fixations per line data. All three variance ratios are indicative of significant differences. However, the ratios evaluating subject variance are much higher than the ratio for samples. Variation in individual differences would seem, then, to be the more important factor in fixations per line performance; the length of the reading period has relatively little influence on performance score. The differences between the mean at i and at subsequent sample periods are very small and not significant until after four hours of reading (VIII). In fact, the significant differences in the *t*-table are few, and occur only when extreme sample periods are compared. The number of fixations per line at VIII represents a six

percent increase (8.22–7.79) over the mean number made during the initial sample period.

Fixations sigma score

Fixations sigma score data are presented in Table IV. The variances assigned to samples and to *S*s are significantly greater than the remainder variance. It might be noted that the ratio of subject variance over sample variance is not significant at the one percent level of confidence, but is so at the five percent level (3.08). Both the individual doing the reading and the length of time a particular individual has read are equally important in determining the variability in number of fixations per line. The differences in the first column of the *t*-table show that the mean sigma score increases in fairly regular order during the reading period. This increment is significant after two hours (IV) of reading, at which time the mean sigma score has increased 12 percent over that at (i) (2.88–2.57). Comparison of subsequent means to the mean at the end of the first half-hour (I) shows a significant increment at two and one-half hours of reading (V), a 10 percent increment over that at I.

Lines read number

The analysis of the lines read number data is shown in Table V. The significance of the first two variance ratios indicates that both the length of the reading time *and* the individual differences of the *S*s doing the reading are important factors in reading speed performance. The difference between subject variance and sample variance (V_2/V_1) is not significant at the one percent level of confidence. The hypothesis is, therefore, still tenable that both time and individuals are equally significant determining factors. Mean performance in number of lines read decreases steadily (except in one instance) and is already significantly different from (i) after one-half hour of reading (I). Mean performance during the second sample period represents a decrement of six percent of the performance at the beginning of the reading period (135.5–143.5). Comparison of subsequent means with the mean at I shows a decrement significant after two and one-half hours of reading.

Regressions number

Regressions number data are presented in Table VI. The variance ratio for samples is significant, but just barely so. The *t*-table shows that the few significant differences in mean performance occur only when performance during (i) is the standard of comparison, which value it will be noted is higher than the general level of the scores at the rest of the sample periods. Mean regression score seems to de-

crease for a while, then to rise suddenly, and then fall again; only the low points of the decrement are significantly different from the performance at (i). The smallest significant difference represents a nine percent decrement from (i) (173.9–191.0). The variance ratios for *Ss* are very high. It would seem that the regression number measure distinguishes among *Ss* regardless of the time in the reading period at which the *Ss* are observed. There is about 20 times as much variation among individual *Ss* as there is among trials with the same individuals. Regressions Number may, then, be regarded as primarily a phenomenon of individual differences. Correlations (r) of the distribution of regression scores at each subsequent sample period with the distribution at (i) (data not shown) are the highest for any of the measures observed in this experiment. This is interpreted to mean that the *Ss* tend to maintain their same relative position during the whole reading period.

Regressions per line

The interpretation of the regression number data (Table VI) may be similarly applied to the regressions per line data presented in Table VII. In order to separate individual differences from the time factor, however, the regressions number measure is better than the regressions per line measure. As in the case of the fixations per line measure, significant decrement in the number of regressions per line does not occur until the last sample period.

The 'per line' measures

Per line measures, such as have been used here, are often regarded in the literature as measures of the efficiency of reading; i.e., how much effort, how many movements must be made to read a line of print. However, these per line measures may be very deceptive when the task of interpreting them is undertaken. If the per line measures had been the only measures analyzed in this experiment, it would have been concluded (except for the results at the last sample period) that little or no significant change in reading efficiency occurred during the prolonged activity. By comparison to the other measures analyzed, the per line measures, then, would be judged poor indicators of possible 'fatigue' effects.

It must be noted, however, that by the very method of their derivation the per line measures may also be statistical artifacts. The measure, Fixations per Line, is derived by dividing the number of fixations made during a sample period by the number of lines read during that sample period; the regressions per line measure is similarly derived, using the number of regresssions as the numerator. It is conceivable that in and of themselves, the per line measures could not

indicate changes during prolonged activity. For example; assume that during a sample period it was found that the Ss on the average made a particular number of fixations per line as derived from a certain number of fixations and a certain number of lines read. But, if during that sample period a larger number of fixations or a smaller number of fixations had been made and a proportionally larger or smaller number of lines read respectively, the number of fixations per line would be the same in all three cases. Therefore, if for several sample periods the number of fixations per line tends to remain the same, all that can be concluded is that either the number of fixations and the number of lines read remained the same or that the number of fixations and the number of lines read varied proportionally. Similar reasoning could be applied in the case of the number of regressions per line. The differential effect of these possibilities on the conclusions drawn in a 'fatigue' experiment is, of course, obvious.

At the last sample period of the present experiment, the number of fixations per line is larger than before, enough so to be regarded as significantly different from the number made during (i). This suggests either that the numerator (fixations) became disproportionately larger (even though it is smaller than the numerator at VII and before) in relation to the denominator, or that the denominator (lines read) became disproportionately smaller (even though it too is smaller than the denominator at VII and before) in relation to the numerator. The graphical representation of mean performance suggests that the ratio of number of fixations to the number of lines read was raised by the number of lines read becoming disproportionately smaller.

The situation at sample period VIII for the number of regressions per line is complicated by the fact that the number of regressions suddenly rises. This increase magnifies the influence of the decrease in the number of lines read. Also, for the first three and one-half hours of reading, the ratio is not as regular as in the case of the number of fixations per line. But neither is the trend in number of regressions regular. Thus, larger or smaller changes in the number of regressions per line does not convey much meaning.

It should be remarked further, in support of the contention that the per line measures may often be statistical artifacts, that the ratio (the per line measures) may be deceptively influenced by the characteristics of the values in either the numerator (Fixations Number or Regressions Number) or denominator (Lines Read Number). A change in a larger number is less apparent than a similar absolute change in a smaller number. In the present instance Fixations Number is the largest number (e.g., for Subject No. 1, Sample i: 1078), Lines Read Number is the next largest (182), and Regressions

Number the smallest (61). Since the number of lines read is the smaller number in absolute magnitude, it would be expected to influence the ratio, Fixations per Line, more as the number of lines read changes than would a corresponding change (i.e., a change of similar absolute magnitude) in the number of fixations. The situation is reversed in the case of the ratio, Regressions per Line, since the number of regressions is now the smaller number. These effects are suggested in the variance ratios summarized in Table VIII. The

TABLE VIII

SUMMARY OF 'PER LINE' AND 'PARENT' VARIANCE RATIOS

	(1) Regressions Number	(2) Regressions per Line	(3) Lines Read Number	(4) Fixations per Line	(5) Fixations Number
V_1/V_3	2.790*	3.571*	13.667*	3.814*	15.248*
V_2/V_3	71.890*	77.632*	45.845*	75.970*	21.439*
V_2/V_1	25.771*	21.737*	3.354	19.920*	1.406

* Significant at the 1 percent level of confidence.

Regressions per Line values (column 2) are more like those for Regressions Number (column 1) than like Lines Read Number (column 3). The effect is not as apparent in the Fixations per Line values; but, if anything, the variance ratios of column 4 are more like those of column 3 than those of column 5.

To avoid the chance of errors of interpretation suggested by the above discussion, and because it is felt that information may be more directly (if not more validly) gained from review of the 'parent' measures, the per line measures will not be considered further.

DISCUSSION

It may be convenient to consider five of the seven eye-movement measures on a sort of continuum based on the characteristics (revealed by the statistical analysis) which these measures may have in common: Fixations Number, Lines Read Number, Fixations Sigma Score, Blinking Number, Regressions Number. At one end of the continuum would be the measure, Fixations Number, which shows both important changes in time and differences due to individuals, but the individual differences are definitely not more effective than the prolonged activity. Lines Read Number also shows important changes in time, but the variance due to individual differences may be more important than in the former measure. Fixation Sigma Score data are similar, but the variance due to *Ss* may be more important than in the case of the Lines Read measure. This trend in the importance of individual differences is continued in the blinking

performance, where it is significantly greater than the variance assigned to time. The regressions number measure is at the other end of the continuum; for the changes in time are slight but the differences due to Ss are very pronounced. The suggested continuum, then, varies from the relatively greater importance of time or prolonged activity to the predominant effect of individual differences.

The number of fixations made and the number of lines read decreased during the four-hour reading period. This decrement is already significant after one-half hour of reading. The standard deviation of the number of fixations per line and the number of blinks made increased during the four-hour reading period; the sigma scores are significantly increased after two hours of reading, and the blinks after one hour of reading. The changes in the number of regressive movements made do not represent a regular decrement or increment, nor are these changes in time relatively important.

There are enough data in the present experiment to demonstrate the significance of both the time variable and individual differences in evaluating prolonged activity. The two factors are at least equally important, but in several instances the factor of individual differences is the more important. The results of the present experiment, then, tend to corroborate the evidence of 'fatigue' experiments in which decrement or change in performance was the primary concern. But the facts that individual differences have been found to be significant considerations and that significant change in performance occurs at different times during the prolonged activity of the several ocular functions argues for regarding fatigue as a phenomenon something more than simply the progressive exhaustion of a peripheral motor mechanism.

At one extreme of the continuum suggested above, when individual differences are relatively minimal in importance (to what they are at the other extreme), significant decrement in performance occurs early in the period of activity. At the other extreme, when subject variance is highly significant, significant change in performance occurs later or not at all. It is also interesting to note that these latter changes are, in the case of blinking and sigma score, towards an increment.

It is not difficult to hypothesize that making fixations or reading lines of print is like doing work; i.e., the action of the extrinsic ocular muscles to move the eyball may be thought of as analogous to the activity of the finger muscles in lifting the weight of an ergograph apparatus. It could be assumed, then, that changes in fixations or number of lines read are measures of the work done by a more or less peripheral neuro-motor mechanism. On the other hand, blinking has frequently been regarded in the literature (**2, 15, 16**) as a possible

measure of 'tension' or stress, or similar psychological phenomena. At least, the rate of blinking has been shown in the experiments cited to vary with the so-called emotionality of a situation, to increase during the performance of a task not necessarily involving the ocular mechanism, and to be more characteristic of the individual than the task.

The nature of regressive movements has been a little more difficult to identify. They may be thought of as related to fixation movements, but in the logic of this measure (that it allows re-fixation of material or 'retracing' of the line of print) is the suggestion that regressive movements may be related in some way to comprehension of the material read. As comprehension becomes poorer, the number of regressions would be expected to increase so as to re-read material not grasped during the first 'reading.'

If the above relationship exists, the fact that regressions did not change as significantly in time as did the other measures requires some explanation. It is possible that the ability to comprehend the material read did not change appreciably—certainly not an unwarranted conclusion considering the evidence of so-called 'mental fatigue' experiments in which performance is maintained at a steady level for long periods of time. Or, assuming that ability to comprehend does decrease during prolonged reading, it may be suggested that some motivational factor breaks down. The *S* does not 'care' after a while about how well he is reading and so makes no effort to keep comprehension at a high level; no increase in the number of regressions made, therefore, would be apparent. There is the further possibility that the reading ability of the present *S*s was so good that variation in comprehension was not great enough to effect regressive behavior. The data of the present experiment, of course, are not adequate to accept one or the other of the above possibilities. It should be mentioned in this connection, however, that no extrinsic motivational device was used in the present experiment except the fee paid the *S*, the instructions to read steadily, and the effect of whatever interest the readers may have had in being the subjects of an experiment. If in further experimentation a valid and convenient means of testing comprehension throughout the reading period itself were used, and the effect of prolonged reading on comprehension scores and on number of regressions were noted, a partial answer to the questions raised would be available.

There is some evidence, however, for regarding regressive movements as related to more central (e.g., comprehension) than peripheral factors. The data for Regressions Number were predominantly influenced by individual differences. It is simpler, then, to correlate this measure with the characteristics of the individuals doing the read-

ing than with the possible effects of prolonged activity. Those characteristics might well be the reading ability or intellectual capacity of the individual readers—obviously more general and central factors than the functioning of a peripheral mechanism. There is also evidence of some periodicity in the data on regressive movements and perhaps also on blinking (see Figs. 4 and 5). After about an hour and one-half of reading and again after three and one-half hours the magnitude of several statistics for these measures was found to be particularly low or high in relation to the same statistics for the other sample periods. Considering these periodic changes, the hypothesis is attractive that, superimposed on the effects of the prolonged activity of the peripheral mechanism, there is a periodic breakdown of the control or determination of the behavior of the ocular functions mediated primarily by central mechanisms.

The above discussion on the possible distinction between fixation and lines read measures on the one hand and blinking and regressive movements on the other also bears upon the interpretation of the fixation sigma score measure. A standard deviation essentially measures the variability of performance—in the present instance variability about the mean number of fixations per line made. This measure has been shown to increase as reading is prolonged, indicating that performance became more variable. The phenomena of variability of performance are much easier to relate to more central factors of control or determination than to intrinsic characteristics of a peripheral organ. On the other hand, the present measure of variability in eye-movement performance is statistically based on the fixation and lines read measures, which latter have been tentatively identified with the work of the peripheral oculomotor system. It supports the general hypotheses of the discussion to find that the fixations sigma score measure reveals, upon analysis, characteristics of both extremes of the continuum which has been suggested. Subject variance may be regarded as not significantly greater than sample variance (as is also true of fixation and lines read measures), but this measure shows an increment (true of the blinking measure) and tends toward establishing the importance of individual differences over time differences (like the regression measure).

All of the previous discussion suggests a rapprochement between what might be called physiological as opposed to psychological theories of fatigue. The former define fatigue in terms or concepts suggesting exhaustion, biochemical anomaly, etc., or posit specific loci of effects and specific causal factors as being responsible for the 'fatigue' of a particular organ system. The psychological theories to explain fatigue prefer either more comprehensive concepts or generalized causal factors, often of central origin, such as motivational and atti-

tudinal factors of control, coordination, direction, etc., of the particular organ system doing work. It is possible that the significant decrements demonstrated in this experiment are instances of specific effects or possible progressive exhaustion of motor units. And it might be that the increments in some measures, the increase in variability, and the importance of subject differences are manifestations of general and perhaps more central factors which have combined their effects with the decremental change occurring in the efficiency of the oculomotor system. The data of the present experiment at least support the conclusion that distinctions can be made in the nature and functions of the measures investigated, whether or not the identifications suggested above are correct.

The *t*-test analyses of the several measures indicate that significant changes in eye-movement performance occurred at various times during the prolonged reading. This would suggest that reading may not profitably be continued beyond a certain length of time if no additional motivation is supplied. However, before this suggestion can assume the importance of a practical recommendation, additional considerations are required. One such is the question of whether the statistically significant differences found in this experiment are significant in the applied or lay-sense of the word. That is, are the magnitudes of the observed changes great enough to interfere with practical efficiency or may reading still be good enough, to all intents and purposes, despite the changes in the component eye-movements of the reading pattern.

In attempting an evaluation of differences for this purpose, *percent* change scores have frequently been used. Usually this score is derived (as was done in the present experiment) by dividing the difference between the means of two sample periods by the mean of the earlier sample period. Though comparison of such scores for the *same* measure may be legitimate in some instances, this technique of expressing gains and decrements as a percentage of the initial score and then using such percent scores to compare *several* measures or functions is open to question. The percentage value so derived has no easily determined standard error, and therefore lacks a valid external criterion wherewith to interpret or judge its significance. It can be readily shown by the data of this experiment that an examination of the percentage change (complete data not shown) is not a reliable method for determining the importance of the change in performance. The number of blinks became significant (at sample period II) when the percent increment increased from 16 to 27. The change in number of fixations was even more significant at I, but the percent change was six. The earliest significant change (IV) in Fixation Sigma Score represented an increment of 12 percent. The number of lines read

was significantly changed at I when the decrement was six percent. The decrement in number of regressions was significant at nine percent. From the above, it will be seen that significant *percent* increment or decrement scores could be 27, 6, 12, or 9. Or reversing the consideration: the percent change in the number of blinks at V was 52; for Fixations Sigma Score at VIII, 19; and for lines read at II, 9. Yet the *t*-ratios of these differences vary but little. At II the percent score for number of blinks was 27 and for Fixation Sigma Score, 8; yet again, the corresponding *t*-scores vary but little.

Returning to the question of practical significance: the mean number of fixations made at the end of the first half-hour (I) is 70 less than the mean number made at (i), or about six percent; comparing fixation performance at the end of three hours of reading with means at (i) and (I), the decrement is 10 and seven percent respectively. At the sample period of significant change (I), the decrement in number of lines read was eight, or six percent of the number read during the first sample period (i); comparing sample periods I and V, 10 less lines were read in the latter. The fixations sigma score at sample period IV increased 12 percent, or 0.32 standard deviation units. During the sample period at which the number of blinks became significantly different from the number made at (i), the actual difference was nine, or less than one-third the number made during the first sample period. When significantly different, the number of regressions was from 18 to 22 (9 percent to 11 percent) less than the number made at (i). Because of the procedure of the present experiment, no definite answer can be given to the question of practical importance which has been raised. It is probable that some of the ocular functions are significantly (in the applied sense) affected while others are not. A tentative conclusion regards the differences more as statistical findings than as practical evidence for limiting continuous reading to short periods of time.

Summary

1. Thirty *S*s read for four hours under relatively normal conditions. Electrical apparatus was used to record the eye-movements of each *S* during the entire four-hour period.

2. Nine five-min. samples were taken of each continuous electroculogram: the first five min. of the four-hour period and the last five min. of each half-hour. For each sample period of reading, seven measures of eye-movements were studied: the number of blinks made in the five-min. interval, the number of fixations, the number of fixations per line, the standard deviation of the number of fixations per line, the number of lines read, the number of regressive movements made, and the number of regressions per line. Analysis of variance and *t*-test techniques were applied to interpret the data of these measures.

3. The number of fixations made and the number of lines read decreased during the four-hour period; this change was significant at the end of the first half-hour of reading. The number of blinking movements made increased—significantly so, after one hour of reading. The standard deviation of the number of fixations per line also increased; this increment was significant after two hours of reading. The changes in the number of regressions made were sporadic and seldom significant. The number of fixations per line and the number of regressions per line did not show significant changes until the end of the four-hour reading period.

4. The analysis of the data indicates that under the conditions of this experiment both time and individual differences are important factors influencing eye-movement performance. In some instances differences between individual *S*s are significantly greater than the changes occurring during prolonged reading.

5. Problems of visual fatigue are discussed in relation to the findings of the present experiment.

(Manuscript received August 13, 1945)

References

1. Anderson, R. L., & Houseman, E. E. *Tables of orthogonal polynomial values extended to N = 104.* Research Bulletin 297. Ames, Iowa: Agricultural Experiment Station, Iowa State College of Agriculture and Mechanic Arts, 1942.
2. Blount, W. P. Studies on the movements of the eyelids of animals. *Quart. J. exp. Physiol.*, 1927–28, 18, 111–125.
3. Clark, B., & Warren, N. A photographic study of reading during a sixty-five-hour vigil. *J. educ. Psychol.*, 1940, 31, 383–390.
4. Dearborn, W. F. The fatigue of the eyes in reading. *Arch. Phil. Psychol., & sci. Methods*, 1906, 4, 123.
5. Fisher, R. A. *The design of experiments.* London: Oliver & Boyd, 1937.
6. Freeman, H. A. *Industrial statistics.* New York: John Wiley & Sons, Inc., 1942.
7. Hoffman, A. C., Wellman, B., & Carmichael, L. A quantitative comparison of the electrical and photographic techniques of eye-movement recording. *J. exp. Psychol.*, 1939, 24, 40–53.
8. Lindquist, E. F. *Statistical analysis in educational research.* Boston: Houghton Mifflin Company, 1940.
9. Lindsley, D. B., & Hunter, W. S. A note on polarity potentials from the human eye. *Proc. nat. Acad. Sci., Wash.*, 1939, 25, 180–183.
10. McFarland, R. A., Knehr, C. A., & Berens, C. The effects of oxygen deprivation on eye-movements in reading. *J. exp. Psychol.*, 1937, 21, 1–25.
11. Miles, W. R. Elevation of the eyeballs in winking. *J. exp. Psychol.*, 1931, 14, 311–332.
12. Miles, W. R. The steady polarity potential of the human eye. *Proc. nat. Acad. Sci., Wash.*, 1939, 25, 25–36.
13. Miles, W. R. Reliability of measurements of the steady polarity potential of the eye. *Proc. nat. Acad. Sci., Wash.*, 1939, 25, 128–137.
14. Mowrer, O. H., Ruch, T. C., & Miller, N. E. The corneo-retinal potential difference as a basis of the galvanometric method of recording eye-movements. *Amer. J. Physiol.*, 1936, 114, 423–428.
15. Ponder, E., & Kennedy, W. P. On the act of blinking. *Quart. J. exp. Physiol.*, 1927–28, 18, 89–110.
16. Telford, C. W., & Thompson, N. Some factors influencing voluntary and reflex eyelid responses. *J. exp. Psychol.*, 1933, 16, 524–539.

THE CONSTANCY OF COLORED OBJECTS IN COLORED ILLUMINATION

BY HANS WALLACH AND ALICE GALLOWAY

Swarthmore College

When a colored object is presented in colored illumination of a different hue, the hue of the light which it reflects will be affected by the illumination. The color of the reflected light will lie somewhere between the hue of the object and the hue of the illumination. If these hues happen to be complementary, the reflected light may, for instance, be of neutral color. The retinal image of the object will, of course, have the color of the reflected light. Yet, the object will be seen more or less in its true hue, and this phenomenon is usually called color constancy. Supposedly there is one condition which must be fulfilled if color constancy is to occur. The colored illumination must be visible on other objects of different hue in the vicinity or, still better, on a neutral background. If the colored illumination is confined to one object or if vision is restricted in equivalent fashion by a reduction screen, the object is seen in a color which is solely dependent on the composition of the reflected light. But when good cues of the illumination are available in the surroundings of the object an object color is perceived which differs from the color of the retinal image in that it is more similar to the hue of the object. It is as if illumination were registered as an independent factor and somehow correction were made 'for the disturbing stimulus variable.'

This formulation of color constancy corresponds to the form in which the other constancies are usually described. The fact that all the constancies lend themselves to analogous formulations has had a profound effect on the scientific treatment which they received. They have been regarded by many as closely related phenomena. Others, notably Koffka, have sought to explain the different constancies separately as consequences of specific functions of their respective sensory fields. An attempt along such lines will here be made regarding color constancy. We shall try to show that the conversion of hue which comes about in constancy situations is identical with the contrast effect which occurs in contrast situations of corresponding structure.

In color contrast, too, the hue seen in a certain area is influenced by the color of its surroundings. Here we have the first resemblance between contrast and color constancy. Our next step will be to show

that the color changes which we encounter in color constancy go in directions which could be predicted from the rules of color contrast The common ground for such considerations should be the colored areas on the retina, for it is when we consider the colors of the retinal images that the problem of color constancy can be appreciated. To start with an example: in green illumination a red object is presented whose color is matched to the illumination to reflect neutral light. Thus the retinal image of the object will be gray and if the object is actually seen on a neutral background its gray image will be surrounded by green, for this is the color of the light which a neutral background in green illumination reflects. If color constancy is effective the object is expected to look more or less red to the observer. However, the same pattern of colors on the retina could come about in a contrast situation, say, in an experiment with colored shadows. A spot of neutral light surrounded by green light on a homogeneous neutral background can produce the same retinal pattern. Here, too, one would expect the neutral spot to look reddish. This case is typical. Even if the red object in green illumination did not happen to reflect neutral, but reddish or greenish light, constancy would still change the color of the object toward red, and a contrast effect of the surrounding green would also tend to modify the hue of the retinal image in the direction of red. The same holds when the color of the illumination is not the complementary but, e.g., blue. In this case the light reflected by the red object would have a purple hue, and contrast would modify a purple area in blue surroundings again in the direction of red, that is, in the direction of constancy. Quite generally, since the light which is reflected when a colored object stands in colored illumination is of a hue which lies between the two colors, and since contrast between this intermediate color and the color of the illumination will tend to enhance their difference, contrast will always tend to change the intermediate color in the direction of the object color.

This fact has, of course, been previously recognized, but no attempt has been made to explain color constancy along these lines. First of all, color constancy and brightness constancy have been regarded as related phenomena, and very strong arguments had been brought forth against an interpretation of brightness constancy in terms of contrast. Secondly, the examples of color contrast with which we are familiar are not striking enough to be compared to the effects of color constancy. Indeed, uncommon observations in the field of color contrast prompted the present attempt. We were greatly impressed by the high saturation of hue which contrast can produce under optimal conditions.

Such observations were made with the following set-up. A projection lantern illuminated a large area of a neutral screen with a colored light except for a circular spot in the center which re-

ceived no light from this source. Neutral light from a second lantern was fitted to fill exactly this dark area. This was done by inserting an opaque slide with a circular hole in this second lantern, while a transparent slide with a circular opaque spot in the center was used in the first lantern. The intensity of the neutral light could be varied by controlling the current of the second lantern, by the use of achromatic filters, or by exchanging the lantern's bulb. The neutral light spot in the center of the colored field assumed the complementary hue and, by adjusting the intensity of the neutral light, conditions for optimal contrast were quickly found. The induced color which was obtained under optimal conditions was fully as saturated as the hue in the surrounding field.

Another casual observation which was made with this set-up reminds one more directly of color constancy: The contrast effect was enhanced when the observer approached the screen, that is when the area under colored illumination filled a larger portion of his visual field. It is well known that the same holds for color constancy. As the area under special illumination becomes larger, constancy is enhanced (Katz's first law of field size).

Our next step was to set up a constancy experiment under comparable conditions. A red color wheel was placed in green illumination. At first the illumination was confined to the wheel by means of the opaque slide with a circular hole. It turned out that a sector 20° blue had to be inserted to let the particular red paper appear neutral in the local green illumination. When the slide was removed and the green illumination covered also the environment, which was neutral gray, the color wheel looked red, but the saturation was by no means greater than that of the red which had been previously produced by contrast using the same green light. As to hue the two shades looked very much alike.

In order to be better able to compare the effects of contrast and constancy, we presented the two situations side by side employing the same colored illumination for both. This again was done with two lanterns. One lantern provided the neutral illumination for the neutral spot which was to show the contrast effect, and the other lantern the green illumination which covered the rest of the background in both situations and the red color wheel in the constancy situation. A transparent slide with a circular opaque spot inserted in this latter lantern again prevented the green light from reaching the neutral spot in the contrast situation. However, before this final arrangement could be presented the two areas which were to show the conversion effects must be matched as to brightness. It is known that color contrast depends on the brightness relationship of the contrast-inducing region and the region which shows the contrast effect.[1] Therefore, we gave the neutral spot which was to display the contrast effect the same brightness that the color wheel had when it was seen in the local green illumination and appeared neutral. This made the brightness relationship between the light reflected by the color wheel and its environment the same as that between the two regions in the contrast situation.

The procedure by which the two areas which are to show the conversion effects are made equally bright is as follows. The green illumination is made to coincide exactly with the color wheel by inserting a slide with a round hole into the lantern which projects the green light. Now the neutral light which issues from the color wheel can be compared directly with the light of the neutral spot. By varying the intensity of the illumination of the neutral spot or by exchanging the screen which forms the background and on which this illumination is visible, the neutral spot can be made to look like the color wheel in the local green illumination. When this is done, the slide can be removed from the lantern which projects the green light and can be replaced by the transparent slide with the opaque spot, so that the green illumination covers in addition to the color wheel the screen, except for the neutral spot, and can produce the two conversion effects.

When the effects of constancy and contrast were compared under these circumstances, no difference could be discovered. The red color wheel under green illumination and the neutral spot in the green en-

[1] Optimal contrast is supposedly obtained when the two regions are of equal brightness.

vironment looked exactly alike as to hue and saturation. They both happened to be of a fairly saturated dark red color.

The same experiment was done with different brightness relationship between the green environment and the areas which show the conversion effects. The same red color wheel was used, but the background screen which previously was of a middle gray color was changed into a light gray. This made the green environment considerably brighter. The brightness of the neutral spot was again matched to the color wheel in local green illumination. (The neutral illumination had to be reduced to compensate for the lighter screen on which the neutral spot was now projected.) When the green illumination covered the whole field except for the neutral spot, it was again apparent that the color wheel and the neutral spot looked exactly alike. Again they were dark red, but compared with the previous experiment they were slightly less saturated. When the experiment was repeated with a dark gray screen, the color wheel and the neutral spot were brighter and still less saturated, but the changed brightness relationship to their environment again affected both in the same manner; they looked exactly alike.

A similar result was obtained when these experiments were performed with a blue color wheel and yellow illumination. When the yellow light was presented on a white screen, color wheel and neutral spot had the same well-saturated blue color. When the middle gray screen was used the saturation of the blue was lower, and with a dark gray screen both the color wheel and the 'infield' in the contrast situation were nearly neutral.

Every experiment was performed with 5 observers who were not acquainted with the significance of the experiment. Each observer first matched the illumination of the neutral spot with the color wheel in local chromatic illumination, i.e., when they both appeared neutral. When this was accomplished, the general chromatic illumination was introduced and the two conversion effects were compared. Occasionally an observer reported a slight difference in saturation of the induced color, but the difference favored as often the neutral spot as the color wheel. These differences were probably the result of small errors in the initial match which was not easy to achieve.

In these experiments the colored region which surrounded the areas that showed the conversion effects had more or less the appearance of a surface under colored illumination, and the induced color in the color wheel looked like a surface color. Thus the phenomenal characteristics which presumably accompany color constancy were present. But even in this respect the neutral spot (contrast situation) did not differ from the color wheel; it also had the appearance of surface color. This is interesting in view of the fact that the neutral spot when it was presented separately, i.e., without illumination in its environment, did not look like a surface color at all, but dimly luminous. And the same was true of the color wheel: When the colored illumination was

confined to it, the wheel looked like a dimly luminous object, and only when the illumination covered also the surrounding screen did the wheel assume the appearance of surface color. Thus, the surface color character came about together with either conversion effect.[2]

In general, it seems to be impossible to detect any difference between the psychological effects of a color constancy situation and of a contrast situation of corresponding structure. One might conclude that as a function of the nervous system color constancy and color contrast are identical phenomena.

There is, however, an objection which could be raised at this point. It is connected with the fact that in the contrast situation, too, the color in the environment was seen to belong to the illumination, at least to the same degree as in the constancy situation, since both conversion effects were brought about by the same environment. It might not be essential that in one situation a colored object in colored illumination is objectively given and in the other a neutral light on a neutral screen; rather, the decisive condition may be that the inducing region has in both situations the same phenomenal character of an illumination. In other words, one might object that essentially we dealt with color constancy in both parts of the experimental situation. According to this view our hypothesis that color constancy and contrast are identical phenomena could only be tested by changing the mode of appearance of the hue in the inducing region, in order to see whether this made any difference in the conversion effects. If color constancy exists as a separate phenomenon, the conversion effect should be greater when the effective hue is experienced as belonging to an illumination (which covers also the area in which the conversion effect is produced) than when it is experienced as the object color of the environment.

This possibility was tested with a somewhat different set-up. Because it lends itself to a more concise experiment we used colored transparency instead of colored illumination. When a colored object is seen through a colored filter or episcotister of different hue, the color of the retinal image of the object is affected by this hue, just as it is affected by the hue of a colored illumination, and the fact that through a colored transparency an object is perceived more or less in its true color poses exactly the same problem as the case of the colored object in colored illumination. Indeed, the hues perceived through colored filters or episcotisters, inasmuch as they were different from the stimulus color, have always been considered products of color constancy.

Again we offered a constancy situation and a contrast situation for direct comparison. The new constancy situation consisted of a narrow red ring ½ inch wide pasted on a middle gray cardboard and an episcotister of bluish green paper which could spin in front of the cardboard at a distance of one or two inches. For reasons which will later become obvious the ring was concentric with the episcotister. The episcotister had four sectors of 50° width, while through its apertures four sectors of 40° of the gray cardboard and the corresponding sections of the red ring were visible. It had previously been found that a color wheel mixture of four times 50° of the green paper and four times 40° of the red paper looked neutral gray. Thus, apart from any conversion effects, the red ring should have looked gray when the episcotister was spinning. Indeed, a ring of the same measurements arranged concentrically on a gray wheel and consisting of four 50°

[2] The significance of this observation will be discussed in another paper.

sections of the green paper and four 40° sections of the red paper looked gray when the wheel was spinning. In the constancy situation, however, the ring was seen in orange-red color behind the rotating episcotister, which had the appearance of a green transparent sheet.

The contrast situation was objectively very similar to the constancy situation. The only difference was that here the gray cardboard with the red ring was arranged as a color wheel (the ring, of course, in concentric position) and that the episcotister was pasted on top of it all. Thus, the whole set-up consisted of one color wheel which was in the greater part composed of gray and green sectors and contained a narrow ring-shaped region (*T* in Fig. 1) which was made up of red and green sectors. When the wheel was rotated this arrangement was equivalent to a ring-shaped neutral region which bordered on both sides on a desaturated green. A contrast effect of this green was expected to make the ring appear reddish.

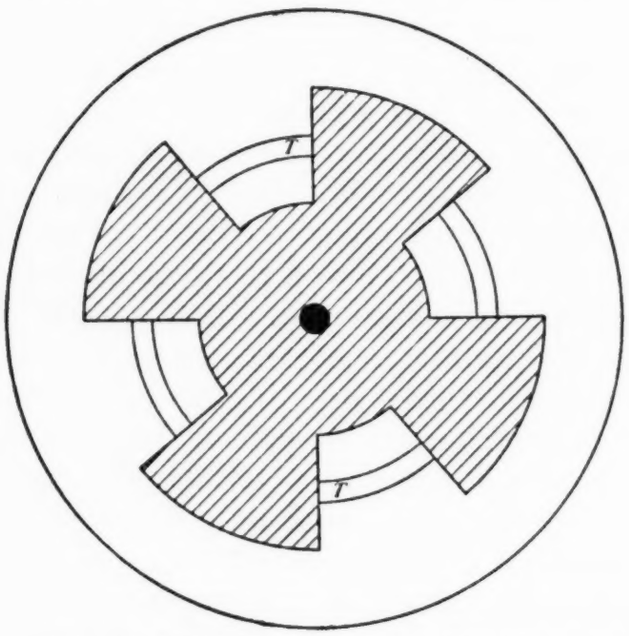

FIG. 1. The hatched region is green, the ring sectors are red, and the remaining area is gray.

Since this set-up was in all measurements identical with the set-up in the constancy situation, it produced exactly the same retinal colors as the constancy set-up. The important difference was that in the constancy situation the episcotister was at a distance in front of the gray cardboard and the red ring and that consequently the green was seen as a detached transparent sheet which seemed to cover the ring and the rest of the field equally, while in the contrast situation the green was seen in the same plane with the ring and did not seem to cover it. In spite of this distinct difference in the mode of appearance of the green, the red ring had in the contrast situation exactly the same hue and saturation as in the constancy situation. Its bright red was an impressive example of a conversion effect under favorable conditions.

The experiment was repeated with a yellow ring and a blue episcotister. The result was the same. The ring showed the same yellow color in both situations, although here, on account of a less favorable brightness relationship between the ring and its environment, the saturation was lower than in the case of the red ring. The

two experiments indicate that the mode of appearance of the effective hue has no influence on the conversion effect.

However, the experiments do not concern the nature of the transparency phenomenon. They merely indicate that the presence or absence of transparency has no influence on the perceived hues and their saturation. The latter depend merely on the colors of the retinal images and on the interaction effects which the color processes following from the retinal colors have on each other. In so far as hues are concerned, the well-known rules of contrast seem to predict the interaction effects in satisfactory manner. The brightness relationship between the region of the effective hue and the region which shows the conversion effect has a strong influence on the saturation of the induced color. This influence probably accounts for the results of earlier experiments which indicated a functional difference between color constancy and contrast.

It is difficult to transform a constancy situation into a comparable contrast situation without changing the brightness relationship between the region of the effective hue and the region which shows the conversion effect, and as long as one is not aware of the strong influence of this brightness relationship one is only too likely to allow such a change to take place.

Koffka (1, pp. 340–341) reports an experiment with colored shadows in which a change from illumination to surface color in the inducing region made the conversion effect smaller. On a white screen illuminated by diffuse white light and a yellow lamp a shadow was cast. This region reflected only white light and looked blue. When the screen, with the exception of the shadow, was then covered with yellow paper so that the paper surrounded the shadow completely, the shadow lost most of its color. Koffka observed that this might be due to the changed brightness relationship. He therefore substituted for the white screen on which the shadow was cast "a gray one of such a value that the shadow on it was of the same brightness as the yellow paper." In spite of the fact that now conditions for contrast became supposedly optimal adding the yellow paper again made the conversion effect smaller.

We reproduced this experiment with simultaneous presentation of the two situations and repeated it many times with wide variations in the brightness relationship, that is with screens of different gray and with different ratios of the white and colored illumination. We were unable to confirm the implications of Koffka's results. Although it was possible to arrange the brightness so that the conversion effect was weaker where the yellow paper was added, we found that much more frequently introducing the yellow paper, which amounts to a higher saturation of the effective hue, leads to a greater saturation of the blue. The same result was obtained with other colors. It must be assumed that, in spite of his attempt to rule out such an influence, Koffka's result is to be attributed to the particular brightness relationship he happened to employ and not to a difference in the mode of appearance of the inducing color.

This paper does not concern itself with the appearance of achromatic colors in chromatic illumination nor with the part that adaptation plays in natural situations in keeping colors in chromatic illumination constant.[3] We have made sure that the conversions which were

[3] Thus, Helson's very interesting work (this *Journal*, 1938, 23, 439–476 and 1940, 26, 1–27) cannot be discussed here. His subjects made their surprising observations after 15 min. adaptation to a very strongly chromatic illumination. It seems likely that adaptation played a larger role in bringing about his results than interaction.

observed were the effects of interaction rather than of adaptation by performing control experiments with strict fixation or with short exposure.

Summary

The effects of color constancy and of color contrast were compared and were found to be identical. In every experiment a situation yielding a constancy effect was presented for simultaneous comparison with a contrast situation of a geometrically similar arrangement of colors. When the colors in the two situations were so chosen that the lights reflected by corresponding regions were equal as to brightness and hue, the constancy and the contrast effects were found to be equal in every respect. It turned out that the mode of appearance of the effective color by which constancy was supposed to be distinguishable from contrast was without influence. Only the quality of the reflected lights and their geometrical arrangement seemed to matter.

(Manuscript received July 7, 1945)

Reference

1. Koffka, K. Some remarks on the theory of color constancy. *Psychol. Forsch.*, 1932, 16, 329–354.

THE WIDTH OF THE BASILAR MEMBRANE IN THE GUINEA PIG[1]

BY SEYMOUR PERLIN, A/S USNR

Princeton University

The characteristics of the basilar membrane have been a matter concern in the various theories of hearing, and especially in the resonance theories, which assume differential tuning by variation in the physical dimensions. Here the width, tension, and loading of the membrane are the chief variables, and of these the width is generally regarded as primary. Yet only recently has much work been done in the refinement of methods for its measurement, and only a few animals of a few species have been studied.

This paper is concerned with the width of the basilar membrane in the guinea pig (*Cavia cobaya*). This animal has been used extensively in auditory research, especially in studies of sound injury and electrical response, and consequently information on the detailed anatomy of its cochlea assumes special importance.

Guild (**1**), using the guinea pig, was the first to show that the width of the basilar membrane varies greatly within a species, and also, contrary to prevailing descriptions, it reaches its maximum about a turn before the apical end is reached. He made three graphic reconstructions of the form of the cochlea, and used these to determine general positions for eight radial measurements,[2] one in each half-turn, in a large number of ears.

An account of the development of methodology, from Hensen and Retzius to Guild, has been made by Wever (**2**). He employed Guild's graphic reconstruction method (**3**) in the measurement of the width of the basilar membrane in the human ear, and adapted it to the measurement of oblique sections. In man, as in the guinea pig, great individual variability was found. On the average, the difference between maximal and minimal widths was six and one-quarter times, and the maximum appeared about one-half turn below the apex.

[1] From the Princeton Psychological Laboratory. Acknowledgment is made to Dr. E. G. Wever for his very generous guidance and assistance and to Lieutenant John A. Guy for his coöperation.

[2] Radial positions are those in which the plane of section passes through the center of curvature of the portion of basilar membrane concerned; this happens at but one place in each half-turn. All other positions are oblique, up to the point where the membrane is cut only on a tangent.

Method and Procedure

This experiment used serial sections of the guinea pig cochlea, with every fifth section mounted. The sections, 20 micra thick in all cases, were made in the vertical plane and parallel to the mid-modiolar axis of the cochlea. Studies were made of 10 cochleas, chosen with respect to histologic preparation, orientation, and preservation. No hearing tests had been made before death.

The weights of the guinea pigs ranged from 210 gm. to 353 gm., with an average weight of 282.5 gm. The lengths of the membrane varied between 13.8 mm. and 18.0 mm., with an average length of 16.1 mm.

Measurements were made at 24 to 26 positions, depending on the form of the cochlea at the apical and basal ends. For every half turn there were two oblique measurements and one radial measurement. The positions of the measurements were located for each cochlea according to the graphic reconstruction method described by Guild.

In this method, a scaled ortho-projection of the line of junctions of the pillar cells of the organ of Corti is obtained. An outline of the procedure follows:

Junctions of the pillar cells are cut tangentially once in each half-turn. Such tangentially-cut sections are plotted on cross-section paper on an axis which is perpendicular to the plane of section and at intervals on this axis determined by order and thickness of sections. The tangential point of the first half-turn is connected to that of the second, which is connected to that of the third, etc., by semicircles, thus forming an approximate spiral which represents the middle portion of the cochlea.

Additional measurements are necessary for the upper apical and lower basal turns. A screw-micrometer ocular is used to measure the distance between the apical end and the lower apical turn appearing in the same section. This distance determines the arc which is drawn from the previously plotted apical turn to the laterally-located apical end. Similarly, the distance between the radial position of the first or lower basal half-turn and a position in the upper basal half-turn in the same section is measured. An arc whose center is on the perpendicular axis then connects the plotted upper basal and the lower basal position. The representation of the extreme basal end is less accurate than that of the remainder, and a certain amount of personal judgment of the general shape is involved.

A calibrated screw-micrometer ocular was used for measurement of the width of the basilar membrane, which was taken as the distance between the spiral ligament and the junction at the spiral lamina. The criterion of judgment as to the end-points of the membrane was the apparent mobility.

Radial positions were read directly. For oblique sections another procedure is necessary (2). In finding the width of oblique sections, the final measurement is determined indirectly by a geometrical construction from the observed distance. By averaging oblique-section measurements on either side of the trangential section, and at approximately equal distances, the width of the membrane at the tangent may also be determined. Thus a total number of 30 to 32 measurements is obtained for each membrane.

It has been computed that the neglect of the height of the spiral reduces all linear distances by about one-third of one percent. Other errors are the element of judgment in determining the attachments of the membrane, and the difference between the semi-circular projection and the true form of the cochlea.

Results

The results of the measurements are shown in Figs. 1 and 2. Fig. 1 is a scaled representation of each of the 10 basilar membranes. In this figure, the widths of the membranes are marked off vertically, and correlate with distances from the basal end indicated on the abscissa. All distances are in mm., but it is to be noted that the width is exaggerated ten-fold with respect to the length.

The total length is not always indicated, since the last measurable width may occur at a point before the end of the membrane. Also,

reliable measurements could not be made at the extreme basal end where the membrane twists in an irregular fashion. To show correctly the length from the basal end, the early portion of the membrane

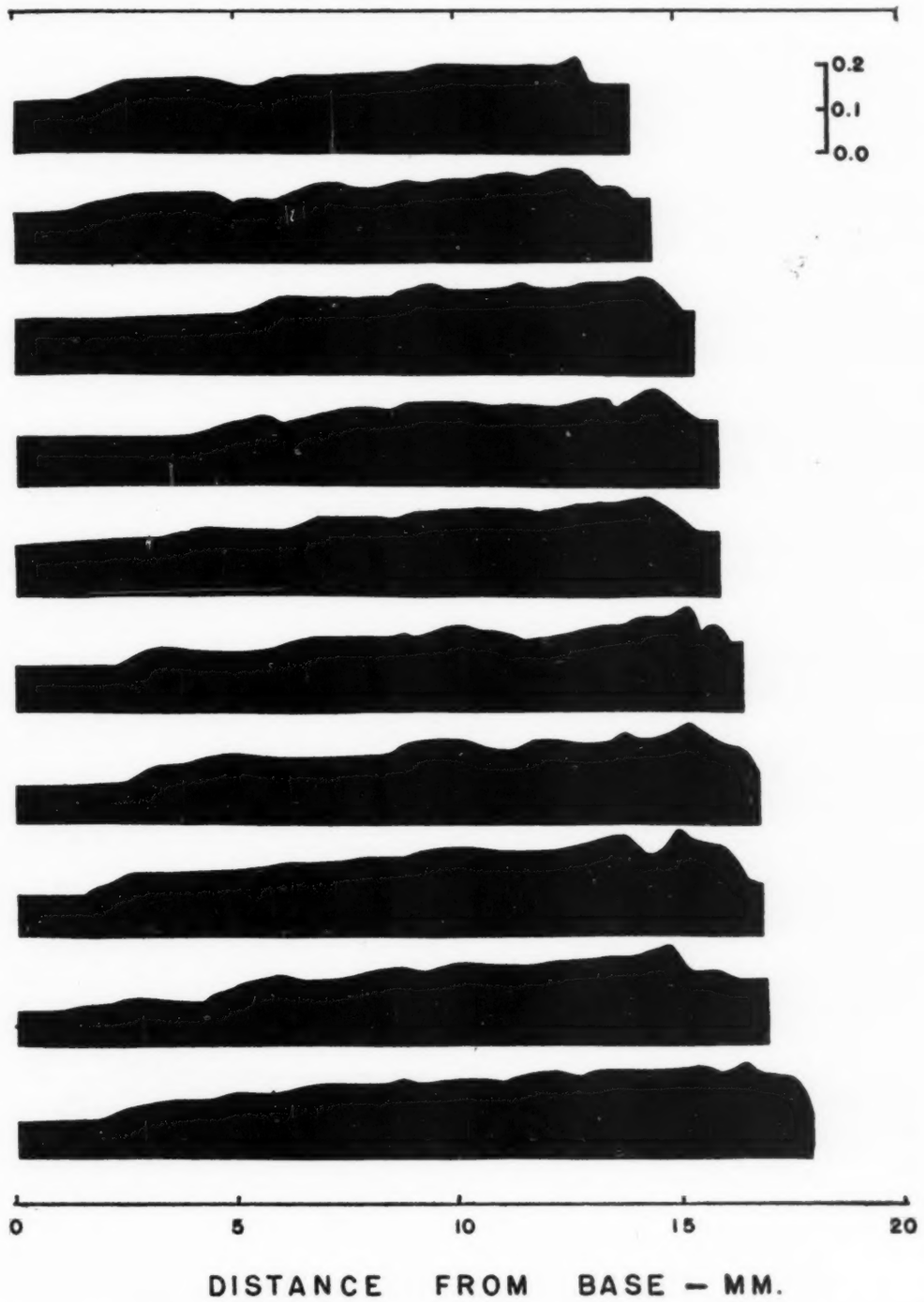

FIG. 1. Measurements of width on 10 basilar membranes. Widths are shown vertically in respect to distance from the basal end. All distances in mm.; but the scale of width is 10 times as great as that of length.

is drawn as of uniform width, but the first measurement occurred only at the upper end of the flat portion.

These measurements show in general an increase in width from the

basal end upwards. However, the following variations occur: there is a rapidly increasing change in width before the third half-turn; a deceleration in the rate of change (change in width per units traversed) in the approach to the third half-turn; a slight spurt afterwards, followed by a slow but definite progression in width; another decrease in the rate of change in the approach to the fifth half-turn, followed by a rapid increase to the maximum. The few measurements possible after the maximum point, which is usually reached three-quarters to a full turn before the end of the membrane, indicate a rapid decline.

The averaging of data of this sort presents particular difficulties on account of the wide individual variations. Three different methods were tried.

The first and most obvious method is one in which the measurements obtained at the various positions—radial, oblique, and tangential—are averaged directly. However, the deviation of the various measurements from this average is relatively large. It appears that the size of the deviations is in part a function of the method itself, for the points being averaged are located by the form of the membrane and not by any determinate distance from the basal end. The form of the membranes varies considerably, especially in the basal turns; and in different membranes, the same radial measurement (e.g., at the second half-turn) differs significantly as regards the distance from the basal end. It is evident that this distance may be the important factor in width determination.

Therefore a second method of averaging was introduced. By interpolation between the observed measurements, a set of readings was obtained for various arbitrary distances along the membrane, measured from the basal end. The average of these readings is represented in Fig. 2.

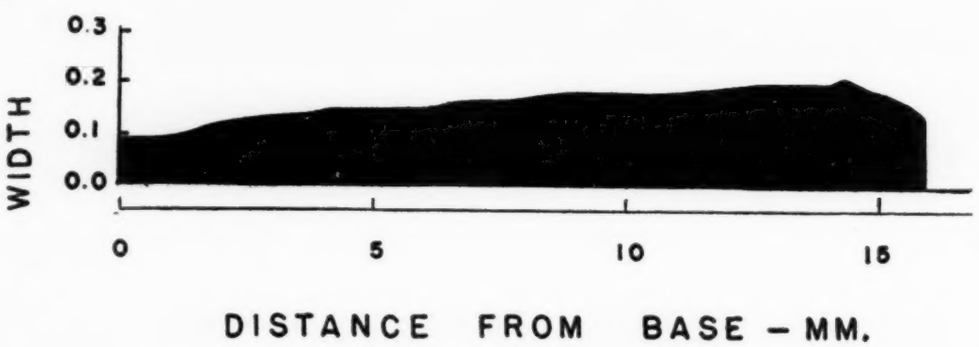

FIG. 2. An average of the 10 membranes of Fig. 1

Owing to the fact that the membranes are of different lengths, this method brings out a further difficulty. The pre-maximum widths of one membrane may be averaged with the post-maximum widths of another. Consequently, after the maximum of the first membrane is reached, the averages are not very meaningful.

A study of the individual membranes and their deviations from this average, does however, indicate the following features. There is a 'beginning' portion in which deviations are considerable, but with the width increasing fairly rapidly in each of the membranes. There is also an 'end' portion, which starts shortly before the maximum point and there, and even more so after the maximum point, shows marked deviations. A definite and hurried decline occurs after the maximum point has been reached. The 'middle' portion, however, shows surprisingly less deviation. For the purpose of separating if possible, or studying in greater detail, these 'beginning,' 'middle,' and 'end' portions, a third method was introduced.

In this last method, a reading was obtained at five percent intervals along each of the 10 membranes by interpolation between the observed measurements, and these readings were averaged. What this method reveals is the average form of the membranes irrespective of their individual lengths. The results are given in Table I. The

TABLE I

AVERAGE WIDTHS OF 10 BASILAR MEMBRANES AT FIVE PERCENT INTERVALS ALONG THE MEMBRANES, AND CORRESPONDING AVERAGE DEVIATIONS

Percent of Length	Average Width in mm.	Average Deviation
5	0.089	0.0102
10	0.105	0.0123
15	0.123	0.0136
20	0.136	0.0129
25	0.142	0.0128
30	0.150	0.0093
35	0.155	0.0045
40	0.161	0.0083
45	0.170	0.0060
50	0.174	0.0045
55	0.181	0.0057
60	0.188	0.0078
65	0.189	0.0054
70	0.191	0.0058
75	0.196	0.0067
80	0.202	0.0057
85	0.208	0.0063
90	0.215	0.0128
95	0.187	0.0121
100	0.150	0.0106

table presents the average widths and the average deviations at the various centile distances.

As the table shows, the membranes are very much alike in the middle region. Between the 25 and 85 centile distances, the deviations are small, as has already been observed in Fig. 1 for the absolute-distance measurements. The earlier and later regions still show large deviations.

An examination of the results showed no significant variation that could be correlated with the body weight of the animal.

Discussion

As stated in the introduction, the basilar membrane figures prominently in many theories of hearing, especially place theories, in the differential tuning of the cochlea. The variations in width are of course most important, and in addition to the previous discussion of deviation in various sections of the membrane, the extent of range demands special consideration and attention.

For the extreme basal portion of the membrane, Guild's (1) measurements indicate an average of 0.062 mm., with a range of 0.049 to 0.073 mm. In the present experiment, as mentioned previously, satisfactory measurements could not be made for this portion of the membrane. The first complete set of measurements was obtained at the oblique section just before the middle of the first half-turn, where the average reading was 0.100 mm.

Guild in his report does not determine an average of his observations of maximum width, but he gives the average of the radial sections at the seventh half-turn where many of the maxima occur. This average is 0.209 mm., with a range of 0.192 to 0.228 mm. Thus, these measurements show a range from basal to apical regions of approximately $3\frac{1}{3}$.

In the present experiment the approximation of a true maximum for each membrane is made more likely by the added oblique measurements. The maximum values found in this manner are somewhat larger than Guild's, and average 0.228 mm., with a range of 0.213 to 0.240 mm. Using Guild's figure of 0.062 as an average minimum and 0.228 as the average maximum, the average variation in width becomes $3\frac{2}{3}$.

Summary

The width of the basilar membrane was measured in 10 guinea pig cochleas, using as a reference point the distance from the basal end.

The measurements indicate a rapidly increasing but widely deviating 'beginning' portion, followed by a deceleration in rate of change before the third half-turn; a slight spurt after the third half-turn; a 'middle' portion of slight but definite progression with relatively little deviation; another deceleration in the approach to the fifth half-turn; a rapid and deviating increase to the maximum point; and an even greater deviation and hurried decline after the maximum point, which was reached about three-fourths to a full turn before the apical end.

The average value of the maximum points measured was 228 micra. Using Guild's figure of 0.062 mm. as the average minimum for the extreme basal portion, there is a range of $3\frac{2}{3}$ times in width.

Theories of hearing which place emphasis on the width of the mem-

brane as a differentiating mechanism must allow for the many variations indicated.

(Manuscript received July 18, 1945)

References

1. Guild, S. R. The width of the basilar membrane. *Science*, 1927, **65**, 67–69.
2. Wever, E. G. The width of the basilar membrane in man. *Ann. Otol. etc., St. Louis*, 1938, **47**, 37–48.
3. Guild, S. R. A graphic reconstruction method for the study of the organ of Corti. *Anat. Rec.*, 1921, **22**, 141–157.
4. Hardy, M. The length of the organ of Corti in man. *Amer. J. Anat.*, 1938, **63**, 291–311.

THE RELATION BETWEEN FREQUENCY OF BLINKING AND EFFORT EXPENDED IN MENTAL WORK

BY M. E. BITTERMAN AND E. SOLOWAY

Cornell University

Luckiesh and Moss (**8**) have placed a good deal of emphasis upon the value of frequency of blinking as an index of the efficiency of conditions of visual work. They have contended that the rate of this involuntary activity of the eyelids increases with the difficulty of the task and reflects the degree of effort expended by the *S* in an attempt to maintain an adequate level of performance. A variety of recent experiments (**2**, **9**, **10**, **11**) have failed to support this conclusion.

It is still possible, however, that there is a tendency for blinking to increase with effort in so-called mental work. A number of studies (**4**, **5**, **6**, **8**) have shown that muscular tension varies with the difficulty of the task, and the work of Ponder and Kennedy (**13**) suggests that blinking may be facilitated by these tensions. Litinski (**7**), although failing to obtain consistent results with blinking as a measure of visual efficiency, concluded that the index might reflect general fatigue, a conclusion which is significant in this connection if fatigue is regarded as a state requiring increased expenditure of effort per unit of work accomplished (**1**). Clites (**3**) has demonstrated that blinking increases (as compared with rest) during mental work of a non-visual character, and his data indicate that the difficulty of the task may be a significant determinant.

If there is a tendency for the frequency of blinking to increase with effort expended in mental work it becomes necessary to account in some way for the negative results obtained in the majority of visual experiments. This problem may be solved by postulating a simultaneously operating counter-tendency. It is well known that the rate of blinking decreases at the onset of visual work (**12**), probably to facilitate vision, and it is possible that the inhibitory tendency becomes more marked as the conditions of seeing become more severe. From these assumptions it would follow that tasks of unequal difficulty might not be differentiated in terms of blinking when the difference in difficulty stems from the visual conditions of work, but that they might be so differentiated when the difference in difficulty stems from other factors. The present experiment was devised to test this possibility.

Procedure.—The plan of experiment was similar to that outlined in a previous communication from this laboratory dealing with the effect of type size on frequency of blinking.[1] In the

[1] For details, see **2**, pp. 283–285.

present experiment the variable was distraction produced by the continuous playing of a record of a man's voice reciting groups of numbers from one to ten digits in length. The intensity level was approximately 60 db. Each of the 10 *Ss* worked for four 15-minute periods on reproductions of the Minnesota Vocational Test for Clerical Workers under illumination measured as 10 foot-candles at the working surface. The record was played during two of these periods and complete silence was maintained during the others. The method controlled practice, fatigue, adaptation to the level of illumination, temperature, and humidity. The procedure of obtaining performance scores and continuous electromyographic tracings of cardiac and eyelid activity was the same as in the previous experiment.

Results.—Work done on the clerical test was computed by scoring the blanks right-minus-wrong, and in addition the percentage of accuracy for each period of work was computed. Table I demon-

TABLE I

THE EFFECT OF DISTRACTION UPON PERFORMANCE

	Distraction	Silence	Difference	Critical Ratio
Output	480.1	467.9	12.2	1.2
Percentage of Accuracy	97.3	97.6	.3	1.3

strates, as might have been expected, that the distraction had no significant effect upon either of these measures. Heart rate, on the other hand, was reliably higher during work than during rest, and higher under conditions of distraction than under conditions of silence. The data are presented in Table II. Table III summarizes the re-

TABLE II

MEAN HEART RATES DURING REST AND WORK

	Mean Heart Rate		Critical Ratios
Rest	80.2	Rest vs. Work (Silence)	6.4*
Work (Silence)	85.9	Rest vs. Work (Distraction)	11.1*
Work (Distraction)	88.0	Work (Silence vs. Distraction)	3.3*

* Significant at the one percent level.

TABLE III

MEAN FREQUENCY OF BLINKING IN RELATION TO DISTRACTION

	Successive 5-Minute Periods			Total
	1	2	3	
Silence	22.7	24.7	22.1	69.5
Distraction	23.2	23.9	24.2	71.3
Difference	.5	−.8	2.1	1.8
Critical Ratio	.3	.8	.6	.5

sults for frequency of blinking. There was no tendency for this index to increase during work under either condition or to be higher under conditions of distraction.

Discussion.—Although there was no significant difference in performance under the two conditions of work, the results for heart rate justify the assumption that the distraction employed was sufficient to necessitate increased effort on the part of the Ss. The negative results for blink rate indicate, therefore, that this index bears no relation to the difficulty of the task. Since it has generally been found in experiments of this kind that distraction is accompanied by increases in muscular tension, it must be concluded that there is no tendency for these tensions to influence frequency of blinking.

The relation of the present findings to those of a previous experiment on type size (2) should be mentioned. In that experiment it was found that performance on the clerical tests varied with type size, although neither heart rate nor frequency of blinking was affected to any significant extent. It might be argued that such results cannot lead to any conclusion with respect to the validity of blink rate as a measure of effort because it may be that the lower performance scores achieved with smaller type indicated that effort did not increase. The present results remove the basis for this argument. Admitting that the conditions of work under distraction were more difficult, the fact that output remained constant must lead to the inference that greater effort was expended in the work. The rate of blinking, nevertheless, did not change.

(Manuscript received July 17, 1945)

Bibliography

1. Bitterman, M. E. Fatigue defined as reduced efficiency. *Amer. J. Psychol.*, 1944, 57, 569–573.
2. Bitterman, M. E. Heart rate and frequency of blinking as indices of visual efficiency. *J. exp. Psychol.*, 1945, 35, 279–292.
3. Clites, M. S. Certain somatic activities in relation to successful problem solving. *J. exp. Psychol.*, 1935, 18, 708–723.
4. Davis, R. C. The relation of action potentials to difficulty and frustration. *J. exp. Psychol.*, 1938, 23, 141–158.
5. Freeman, G. L., & Lindley, S. B. Two neuro-muscular indices of fatigue. *J. exp. Psychol.*, 1931, 14, 567–605.
6. Hadley, J. M. Some relationships between electrical signs of central and peripheral activity: II. During 'mental work.' *J. exp. Psychol.*, 1941, 28, 53–62.
7. Litinski, G. A. Recording of blinking as a method of studying visual fatigue in children resulting from reading. *Sovet. Vestnik Oftal.*, 1934, 4, 275–286.
8. Luckiesh, M., & Moss, F. K. *Reading as a visual task.* New York: Van Nostrand, 1942.
9. McFarland, R. A., Holway, A. H., & Hurvich, L. M. *Studies of visual fatigue.* Boston: Graduate School of Business Administration, Harvard University, 1942.
10. McNally, H. J. The readability of certain type sizes and forms in sight-saving classes. *Teach. Coll. Contrib. Ed.*, 1943, No. 883.
11. MacPherson, S. J. The effectiveness of lighting—its numerical assessment by methods based on blinking rate. *Trans. illum. Engng. Soc.*, 1943, 38, 520–522.
12. Newhall, S. M. The control of eyelid movement in visual experiments. *Amer. J. Psychol.*, 1932, 44, 555–570.
13. Ponder, E., & Kennedy, W. P. On the act of blinking. *Quart. J. exp. Physiol.*, 1927–28, 18, 89–110.

THE EFFECT OF NEGATIVE INCENTIVES IN SERIAL LEARNING. I. THE SPREAD OF VARIABILITY UNDER ELECTRIC SHOCK

BY G. RAYMOND STONE

Indiana University

Thorndike's (3) generalization that negative incentives do not act in a positive way to eliminate the response they follow, but actually strengthen the response a little, has been supplemented by Muenzinger and Dove's (2) demonstration that, although the negative incentive does not eliminate, it introduces a relative variability of action which is not limited to the punished response alone, but spreads across adjoining rewarded responses with a progressively decreasing effect in either direction from the punished response. The general interpretation of these results may well be that negative incentives function in serial learning only as indirect aids to the selection of correct responses, whose fixation is largely a direct function of positive incentives. The fact that negative incentives have some fixating function gives credence to the point of view which regards inhibition in the learning process as a competition of responses, a competition not only between rewarded responses, but to a lesser extent between rewarded and punished responses.

Together the gradients of reward (fixation) and punishment (variability) overlap in any serial learning situation and would produce, therefore, a maximum of competition in the middle portions of the series. Whether or not the maximum difficulty is the exact middle will be determined by the symmetry of the gradients existing throughout the series, and it is important, therefore, to investigate the nature of both reward and punishment gradients as a function of their serial position in the learning task.

Apparatus and Procedure

A Cornell-Coxe memory drum was modified to present learning items through an aperture in a large plywood screen which separated the *S* from the *E* and the apparatus. Two 39-item lists of two-syllable adjectives were presented to each of 32 college students at a rate of one item every three sec. with no intertrial interval. The *S*s were read instructions as follows:

> You are to be presented with a list of words, each of which is associated with any number or combination of numbers between one and ten. You are to respond to each word immediately with the first number between one and ten inclusive which occurs to you. After each of your responses I will do one of three things: (1) if the number you call out is one of the ones which is associated with the particular word, I may answer 'right'; (2) if your response

is not one of the ones associated with the word, I may say 'wrong'; or (3) I may say nothing at all, in which case you will not know whether you are right or wrong.

We will go through the list several times; each word will be exposed for only three seconds. Although you will not be able to learn the list completely because of the very nature of the experiment, we are interested in your percentage of correct responses. Remember that any one word may have all numbers between one and ten associated with it, or, it may have only one number associated with it. Be sure that you call out some number for every word.

Do you understand these directions?

Before you are through we will go through two different lists and in the learning of one of them there will be an occasional small shock instead of the word 'wrong' for an incorrect response. You will be told which list.

Regardless of the *S*'s responses, the order of incentive responses called for each item in the list was always the same. Using R for the word 'right,' W for the word 'wrong,' and N for no response, this order was as follows: N N W N R W N W R N N N *R R R R W–S* N N N N N N *W–S R R R R* N N N W R W W N N N N. Responses to the underlined items were analyzed under two conditions: (1) when the W–S items represented the word 'wrong,' and (2) when the W–S items represented a short electric shock applied to the *S*'s right wrist through electrodes leading from an inductorium. The shock intensity was adjusted at the beginning of each *S*'s session until *S* reported a sharp needle-like prick. No control of shock variation was employed other than the introduction of a 1000 ohm stabilizing resistor in series with the *S*.

The conditions (shock and control) and the lists were completely counterbalanced. Each *S* was run 10 trials under each condition in a one-hour experimental session.

Results and Discussion

The mean number of times that any response to the 10 significant list items was repeated on the next trial has been computed for each condition, and is presented with the significance of the mean differences in Table I. Although the difference between the two conditions

TABLE I

Mean Number of Repetitions of Responses to the List Items Having 'Right' and 'Wrong' as Their Incentive Response. Significance of the Differences

List Response	Condition	Mean	Mean Difference	σ diff.	t	P
right	control	5.66	−.09	.55	.16	.87
	shock	5.75				
right	control	5.47	.97	.51	1.90	.07
	shock	4.50				
right	control	5.03	.53	.48	1.10	.28
	shock	4.50				
right	control	4.63	.13	.56	.45	.66
	shock	4.50				
wrong	control	1.91	−.50	.35	1.43	.17
	shock	2.41				
wrong	control	3.03	.40	.50	.80	.53
	shock	2.63				
right	control	4.68	1.09	.63	1.71	.10
	shock	3.59				
right	control	4.59	.50	.54	.93	.36
	shock	4.09				
right	control	4.50	.66	.50	1.32	.20
	shock	3.84				
right	control	4.09	−.04	.50	.06	.96
	shock	4.13				

is not statistically significant for any one item comparison, the means representing the combined data for the eight 'right' responses in each condition are reliably different ($t = 2.47$; $P < .02$). The four gradients, as measured by the mean difference between repetitions to the first and the fourth right response of the series, are reliable except for the control gradient which followed the wrong response; the mean difference here was .59 ($t = 1.18$; $P < .30$).

When the data are analyzed as a function of practice, the gradients are revealed as early as the second trial although the difference between the conditions, as measured by the repetitions to right responses, does not show up until the fifth trial.

The data are suggestive when compared with Muenzinger and Dove's (2) earlier results, in which these investigators found a bidimensional gradient of variability surrounding a punished (the word 'wrong') response. The order of crucial responses in their study began with a wrong and was followed by seven rights and then another wrong They found that variability of responses decreased across the right responses following the first wrong until it reached a minimum where it increased again progressively until the second wrong. The two gradients they regarded as belonging to the punishment and could therefore be considered bidimensional.

The results obtained under the shock condition of the present study repeat the bidimensional character of variability, but for the control condition a discrepancy exists. Instead of a progressively decreasing variability following the punished response, there is an increasing gradient across the right responses, indicating greater relative fixation of a right response the closer it is to the punished one.

A schematic representation of the gradients surrounding the word 'wrong,' with regard only for the order of right responses in relation to the wrongs, is presented in Fig. 1. The actual serial order of the gradients is here reversed, as the gradient following a wrong occurred first in the Muenzinger and Dove study and last in the present study. The discrepancy is considered, therefore, a function of the serial position of the punished response, even though there were other ways in which the two studies differed; the exact nature of the difference will await further experimentation some of which is now in progress.

The present study varied from that of Muenzinger and Dove by employing items to which the *E* gave no incentive response. The crucial response series was isolated by employing three no-response items before the first gradient series, six between the two gradient series, and three following the second gradient series. The nature of the gradients is thereby not considered a function of the spread of responses other than those studied, unless, of course, no-response

items can be considered to have an incentive function of their own to the *S*.[1]

The general effect of such isolation, supposedly, is to make the crucial series more identifiable to the *S*, and the gradient discrepancy might also be attributed to this factor. That this is probably not the case is suggested by three considerations: (1) one might expect the effect to be increased by increasing the amount of the punishment

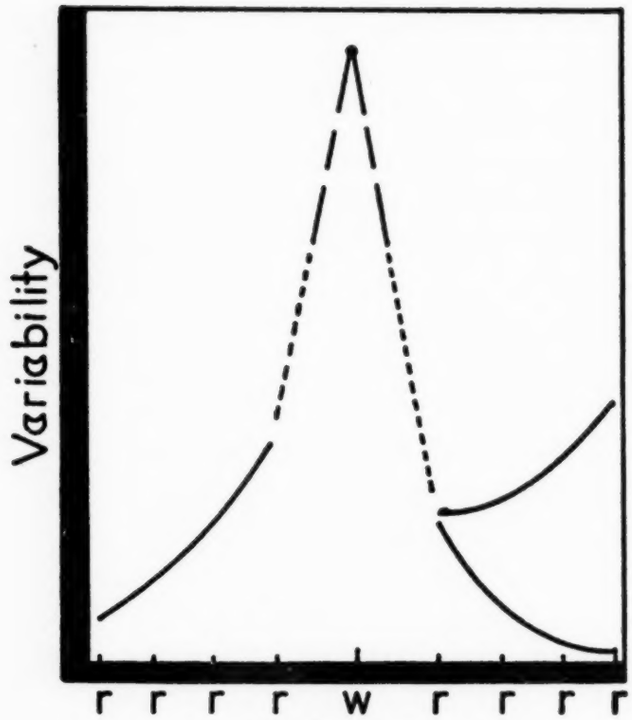

FIG. 1. Schematic gradients of variability surrounding the word 'wrong.' The generalized results of the present study (increasing gradient across 'right' responses after the word 'wrong') are compared with those of Muenzinger and Dove (the decreasing gradient after 'wrong'). For a more detailed explanation, see the text.

(shock), whereas this is not the case; (2) although the percentage of responses was approximately doubled, this did not materially change the form of the initial gradient leading to the punished response; and (3) a similar isolation of rewarded items by memorizing their responses prior to the learning did not distort the gradient of reward (Muenzinger and Dove).

To say that the gradient of variability following a punished response is a function of serial position is equivalent to saying that the effect of punishment is a function of the amount of inhibition or the difficulty of learning. The present results suggest that light punish-

[1] Courts and Waggoner (1), e.g., suggest that the absence of an informative signal may act as a reward. The present data tend not to verify the suggestion.

ment produces less variability in the more difficult portions of learning than in the easy, and that this can be overcome by increasing the amount of punishment in the difficult portions. The relation that this induced variability bears to learning efficiency is problematic, but that it is a function of learning difficulty and the nature of the learning skill (i.e., whether predominantly verbal or motor) seems certain.

Summary

The gradient of variability of responses which surrounds a punished response has been studied as a function of the type of incentive used and the isolability of the punished response. The present study, which is the first of a proposed series on the effect of negative incentives in serial learning, employs slight modifications of the earlier Muenzinger and Dove investigation.

Thirty-two college students were presented 39-unit lists of two-syllable adjectives exposed serially one every three seconds, and were instructed to associate numbers between one and ten, inclusive, with the list units. For each response of the *S* the *E* did one of three things: called out 'right,' called out 'wrong,' or gave no response at all. Although the *S*s were instructed to think of these responses as informative cues, they were always given in the same order regardless of the *S*'s response. Beginning with the tenth list item, the responses were: three no-response, four rights, one wrong, six no-responses, one wrong, four rights, and three no-responses. Under a second condition, a small electric shock was applied to the *S*'s right wrist in place of the word 'wrong.' The results were analyzed in terms of the number of times the *S*'s responses were repeated on successive trials.

1. Positively accelerated gradients of variability exist under both experimental conditions across the four rewarded items which lead to the punished response.

2. The variability gradients across the four rewarded items which follow the punished item are differential in direction according to the experimental condition; the shock condition has a progressively decreasing gradient, while the 'wrong' condition has a progressively increasing gradient.

3. The responses under the shock condition are significantly more variable as a whole than those under the word 'wrong' as punishment.

4. The gradients appear as early as the second trial, but differences between the two conditions do not appear until trial five.

5. The increasing gradient of variability following the word 'wrong' as punishment diverges from the Muenzinger and Dove results taken under similar, but not the same, conditions; the major experimental difference between the two studies is the serial locus of the

post-punishment series, and the discrepancy in results is interpreted as a function of this difference.

6. The general effect of interpolating no-response items before and after the responses to be studied is to increase the percentage of repeated responses in the gradient series, without, however, modifying the form of the gradient.

7. The results of the study are interpreted to support the generalization that light punishment is less effective in producing variability in the more difficult as compared to the easier portions of the learning task; in addition, that increases in punishment severity will overcome the difference. The relation of this variability effect to learning efficiency is thought to have an optimal range probably dependent on the stage of practice and the nature of the task.

(Manuscript received July 23, 1945)

References

1. Courts, F. A., & Waggoner, D. The effect of 'something happening' after a response. *J. exp. Psychol.*, 1938, **22**, 383–387.
2. Muenzinger, K. F., & Dove, C. C. Serial learning: I. Gradients of uniformity and variability produced by success and failure of single responses. *J. gen. Psychol.*, 1937, **16**, 403–413.
3. Thorndike, E. L., et al. *The fundamentals of learning.* New York: Teacher's College, Columbia University, 1932.

GENETIC CHANGES IN SEMANTIC CONDITIONING *

BY BERNARD F. RIESS

Institute for Research in Child Psychology, Hunter College

Introduction

The experiment here reported is essentially an extension to the problem of the genesis of certain language relationships of a technique suggested by Razran (13) and further developed by the author (15) in an earlier paper. The psychological study of semantics was long neglected by psychological discipline in favor of problems more susceptible to objective experimentation and less complicated by experiential variables. In the absence of adequate techniques, speculation took the place of observation and language study was left to the devices of the philologists, who rarely looked upon it as a problem either of social learning or individual habit formation. However, with the discovery and extension of conditioning, studies of verbal patterning and equivalence were made, first by Russian investigators in several laboratories in the U.S.S.R. and then by American psychologists. The general technique employed by these investigators was the study of generalization effects from a conditioned verbal stimulus to words semantically related to the conditioned stimulus. This is the phenomenon now identified as mediated generalization. It was found that generalization or transfer could be established from a word to such relata as the object denoted by the word and to phonetically and meaningfully related words like antonyms and synonyms. The processes conditioned to the verbal stimuli were pupillary and salivary reflexes. A summary of the work of the Soviet psychologists in this field is to be found in the paper by Keller (9).

Recently, Foley, Cofer and their collaborators have published a series of papers using memory methods for the study of mediated generalization and the interpretation of verbal behavior. After discussing the general principles of semantic conditioning and mediated generalization in relation to verbal behavior (1), these authors performed experiments involving the presentation of lists of words for memorization followed by study of the gains in retention from these lists to lists related to the original in various semantic ways. Among the relationships studied were antonyms (2), homophones, first and second order synonyms (4). In other experiments, the authors varied the language background of the *Ss* (6) and the professional and occu-

* This study was made possible by a grant-in-aid from the Social Science Research Council.

pational fields in which the *Ss* were trained (**5**). In the complete experiment as described in these papers mediated generalization in adults proved to vary predictably with language generalizations previously determined by a priori reasoning (**1**).

Using much the same methods, Goodwin, Long and Welch (**8**) studied the mediated generalization of species-genus relationships among children. Training on the lower, species level transferred significantly to the higher or genus level. This is one of the few studies which approach the problem from a genetic standpoint.

Razran's series of extensive researches into relational conditioning cannot be adequately summarized here. It suffices to indicate that he has shown the transfer value of semantic relationships in adults using a conditioned salivary technique with antonymic, synonymic and homophonic stimuli (**13**). In other papers, he has shown the transference of single words and phonetographs such as dark-mark, again using adults as subjects (**11**). Theoretical considerations and the importance of generalized language habits are posited in two other articles (**10**, **14**). It is interesting to note that in the first of these articles (**10**) the author states that the greatest transfer as a result of language was produced by a habit system established in early childhood. This is the earliest reference to the importance of a genetic approach to semantic conditioning.

Diven (**3**), Keller (**9**) and Wylie (**16**) have all studied the semantic effect in conditioning by various methods. The former found a spread of generalization from a word to other meaningful stimuli using GSR as the conditioned process. Keller also used the GSR to demonstrate mediated generalization from a picture to another similar visual representation. Transfer was not found when a verbal stimulus was substituted for the graphic representative. Wylie repeated the early experiments of Razran and Riess with results slightly at variance with those previously found. However, she used electric shock to produce the GSR and hence the procedure was not sufficiently similar to that of Razran and Riess to call for explanation of the divergencies.

The present study differs from the other experiments in two significant ways. It conditions the electro-dermal response (EDR) to a word for which a homophone, an antonym and a synonym exist, and it explores these semantic relationships as they are established in four age groups ranging from the early stages of visual language acquisition to the mature utilization of abstract verbal stimuli. The experiment therefore makes possible the development of a gradient of semantic relationships as a result of mediated generalization. These gradients are then posed as a function of the total experience of the organism in a world of increasing quantitative and qualitative changes in verbal behavior.

METHOD

1. *Subjects.*—The plan of experimentation called for four groups of *Ss* differing only in chronological age. All of the persons used were female and drawn from New York City elementary and secondary schools and from Hunter College. In all groups but that from the College, the IQ's obtained were Stanford-Binet's. No IQ's were available for the college students, but their American Council of Education Psychological Examination percentile scores indicated ranks above the median for the nation-wide norms. Table I shows the mean ages, IQ's and ranges for

TABLE I

AGES AND IQ'S OF THE FOUR EXPERIMENTAL GROUPS

	Group I		Group II		Group III		Group IV	
	Mean	Range	Mean	Range	Mean	Range	Mean	Range
CA	7–9	7–0 to 8–9	10–8	10–0 to 11–8	14–0	13–5 to 14–10	18–6	17–2 to 20–1
IQ	112	92–135	116	98–129	120	101–138	—	—
N	21		22		25		24	

the various groups. All age differences are obviously significant. IQ differences are not statistically reliable.

2. *Stimuli.*—The stimuli for the experiment consisted of five series of words, one of which was exposed only in the pre-conditioning test period. The other groups constituted the major bases of comparison and experimentation. All stimulus words were selected from reading material in use at the educational level from which a particular group of *Ss* was selected. In this way the relative difficulty of the various stimuli for the four age groups was equated.

The words in the series for the preconditioning period combined the 20 items to be used in subsequent sections of the study with 30 neutral words. The latter were not thereafter again presented to the *Ss*. These two groups of words, 20 crucial and 30 neutral, comprise the list hereafter referred to as the pre-test series. A separate pre-test list was evolved for each age group.

The 20 crucial words in the pre-test list were selected so as to demonstrate three types of semantic relationships: homophonic, antonymic and synonymic. Thus there were, in each list of 20 words, five actual stimuli and five homophones, antonyms and synonyms. Table II shows the stimuli in each list of crucial words for the four age groupings.

3. *Apparatus and procedure.*—In general, the apparatus and procedure used in this study were similar to those employed in the previous experiment by the author (15). A simple bucking circuit resistance box was used to measure the electro-dermal response (EDR). Basically the procedure involved three stages. In the first, preliminary determination was made of the extent of the EDR, in microamperes, to each of a series of words presented seriatim (pre-test series) and also to the sound of a moderately loud buzzer. The second stage established a conditioned response to a word from the crucial list by reinforcement with the buzzer. In the third stage, each *S* was tested to determine the amount of transfer of EDR from the conditioned stimulus to each of several words related semantically to the conditioned stimulus. At all times, the words were exposed to each *S* individually by means of lantern slides. The size of the letters in each stimulus was adjusted to the educational level of the *Ss* by appropriate backward and forward movement of the projector.

In the pre-conditioning stage each *S* was instructed to observe the word to be projected on the screen in front of her. She was told that she would be called upon to recall the word at a later date. Warning was also given that buzzers would be sounded during the session and the electrodes were explained as necessary to determine the *S*'s sweat reaction to the temperature of the closed and windowless room.

Stimuli were presented by lantern-slide projection with an exposure time of three sec. and an interstimulus interval of five sec. The order of the words was varied systematically from *S* to *S*. At five points during the session the buzzer was sounded between presentations of the verbal stimuli. After an interval of a week, the *Ss* were again tested with the neutral words

TABLE II

CRUCIAL STIMULI FOR THE EXPERIMENTAL GROUPS

Group	Word	Homophone	Antonym	Synonym
I	WON	one	lost	beat
	WEAK	week	strong	poor
	CLOSE	clothes	open	shut
	FATHER	farther	mother	dad
	WHOLE	hole	part	all
II	SEA	see	land	ocean
	MAID	made	man	girl
	RIGHT	write	wrong	correct
	MALE	mail	female	man
	PALE	pail	dark	light
III	PLAIN	plane	fancy	simple
	BUY	by	sell	purchase
	BOY	buoy	girl	lad
	ATE	eight	starved	fed
	REAL	reel	false	exist
IV	BARE	bear	hide	expose
	COARSE	course	smooth	rough
	MINOR	miner	major	lesser
	FREEZE	frieze	warm	chill
	PEAK	peek	trough	crest

in the pre-test series and asked to recall those previously seen. New words were added to take the place of the crucial stimuli. In this way the validity of the instructions was established. Any words to which the EDR was more than five times that obtained in the preliminary exploration were eliminated as carrying an over-load of emotional coloration.

In the conditioning stage each conditioned stimulus was inserted in a group of neutral words, and the whole list was exposed seriatim with a three-sec. exposure time and a five-sec. interstimulus interval. The order of appearance of each crucial stimulus within the list was varied randomly from trial to trial. The reinforcing buzzer sounded throughout the exposure of the conditioned stimulus and continued for another two and one-half sec. The entire procedure was repeated five times, and then the conditioned stimulus was tested without the buzzer. If the EDR at that test was not at least three times the magnitude of the response to the word prior to reinforcement, the conditioning was resumed. Testing was done every five trials. A conditioned response was held to exist when the EDR of three times the original persisted to three successive exposures of the conditioned stimulus. The amount of the first of these three EDR's was then used in tabulating the results. In all cases, the *S*s were told that they were to try to memorize the list of stimuli and to report when they had done so.

The transfer stage of the procedure then followed. As soon as the conditioned response was obtained, the *S* was shown the test words semantically related to the particular stimulus word used in that part of the experiment. The test words were presented in random order. EDR's were obtained for five successive exposures of the transfer stimuli. Extinction of the CR was then begun by repetition of the conditioned stimulus without reinforcement by the buzzer and the extent of spread of extinction to transfer stimuli noted. A week after completion of each word, the *S* was started on a new series using another conditioned stimulus word. Thus there were, for each *S*, five separate experimental sessions.

RESULTS

The data accumulated in this part of the study are summarized in Tables III and IV. In the former are shown the comparative scorings

TABLE III

MEAN PERCENT AND RATIO OF PERCENT GAIN FOR SEMANTIC WORDS TO PERCENT GAIN FOR CONDITIONED WORDS

Entries in columns of means are based on the mean EDR of each *S* to five words in each stimulus category

Stimuli	Group I		Group II		Group III		Group IV	
	Mean % Gain	Ratio % Gain to CS	Mean % Gain	Ratio % Gain to CS	Mean % Gain	Ratio % Gain to CS	Mean % Gain	Ratio % Gain to CS
Word	221.52	—	226.64	—	242.72	—	281.00	—
Hom.	158.57	71.58	65.18	28.76	59.92	24.69	52.16	18.56
Ant.	139.10	62.79	97.23	42.90	76.96	31.71	103.08	36.68
Syn.	129.43	58.43	59.86	26.41	109.76	45.22	148.50	52.85

TABLE IV

MEAN DIFFERENCES IN PERCENT GAIN, *t*- AND *P*-VALUES

Minus signs before mean differences indicate relative superiority of second member of pair compared

	Stimuli Compared					
	Word-homo.	Word-anto.	Word-syno.	Homo.-anto.	Homo.-syno.	Anto.-syno.
Group I						
Diff. between means	62.95	82.42	92.09	19.47	29.14	9.67
t	6.96	12.14	12.99	2.33	3.39	1.57
P	.01	.01	.01	.02	.01	.20
Group II						
Diff. between means	161.46	129.41	166.78	−32.05	5.22	37.37
t	17.38	13.75	20.05	2.72	.48	3.39
P	.01	.01	.01	.01	.70	.01
Group III						
Diff. between means	182.50	165.76	132.96	−17.04	−49.84	−32.80
t	10.75	11.04	7.19	1.05	2.56	2.17
P	.01	.01	.01	.30	.01	.02
Group IV						
Diff. between means	228.84	177.92	132.50	−50.92	−96.34	−45.42
t	10.61	5.07	6.25	1.56	6.04	1.42
P	.01	.01	.01	.20	.01	.20

of the four age groups on the conditioned stimuli and on the semantically related words. The second table shows the reliabilities of the differences between mean scores for the various age and semantic groups.

Examination of Table III reveals that there was a shift in the transfer of EDR conditioning to various semantic relationships. In the youngest group, the transfer was greatest from the original conditioned stimuli to the homophones of these words. This is indicated

by a mean percentage gain of 158.57 for the homophones as contrasted to percentage gains of 139.10 and 129.43 for antonyms and synonyms, respectively. It should be noted that each mean percentage gain is the result of tabulating the scores of all 21 *S*s in this group on each of five original words, five homophones and similar numbers of antonyms and synonyms. Expressed somewhat differently, the mean gain of 221.52 for the original words in Group I is the mean of a series of 21 means, each of which was derived from five separate entries, each representing the increase in EDR from the original exposure to the first response in the criterion series after conditioning. So also, in the entries under homophones, etc., each mean was derived from a series of means representing five observations each. In all cases, percentage gain refers to the increase in EDR from the original pre-test list to the post-conditioning exposure of the semantically related stimuli.

The homophone, whose superiority is demonstrated in the lowest age group, does not however maintain its place in the other age groups. With an increase in C.A. to 10–8 years, the homophone drops to second place and the antonym receives the greatest amount of transferred EDR. In both younger groups, I and II, the synonym is the least transferred relationship. Group III with a mean chronological age of 14 years shows greatest transfer to the synonym with the antonym in second place. This same trend is repeated in the scores for the oldest group. In summary of this section of the study, we can state that between the ages of eight and ten transfer of EDR shifts from a homophonic to an antonymic relationship. Between the ages of ten and fourteen, a further shift displaces the antonym in favor of the synonym.

The gradient of semantic relata which was just demonstrated in terms of percent gain in conditioning of the EDR is shown again in Table III in the second column for the various age groups. Entries in these arrays show the relationship of the gains in each semantic group to the gains on the original word stimulus. In simpler form, the data here indicate the ratio of gains in homophones, antonyms and synonyms to those in the original conditioned words. Thus, for Group I, the gain in homophone EDR was 71.58 percent of that for the conditioned stimulus, that of the antonym was 62.79 percent, and that of the synonym was 58.43 percent. Again we find the order of relata the same as that above, namely, homophone, antonym, synonym. The other age groups show ratios consistent with the conclusions arrived at above. In general then, the gradient of semantic relationships for the whole group is from homophone through antonym to synonym as one progresses from eight-year olds to college students.

Table IV gives the mean differences, *t*- and *P*-values for these differences for the various age and semantic stimulus groups studied in the experiment. Wherever the differences between the means are prefaced by a minus sign, the second member of the pair is larger. The first half of the table gives the comparisons between the original words conditioned to the EDR and the semantically related transfer stimuli. Differences among the latter are shown in the remaining section of the table.

Study of the first three columns reveals that all differences favor the original stimuli over the semantic relata and are without exception significant at levels far beyond the minimum, .01, reported here. That these comparisons should be so significant statistically is not surprising, since the original word should theoretically carry more conditioned EDR than its mediated generalization transfer. The order of differences within each age group again demonstrates the same relationships reported above; i.e., a gradient of homophone, antonym, synonym for Group I; of antonym, homophone, synonym for Group II; of synonym, antonym and homophone for Groups III and IV.

The reliabilities of differences are more varied within the three arrays covering the inter-comparisons of the antonym, synonym and homophone stimuli. When the differences and their reliabilities are ranked in accordance with the gradients described above for each age category, it will be noted, that with five exceptions, all differences are significant at the .01 level. In each age group, the differences between the second and third members of the triad gradients are significant at levels of .10 to .20 (Group I), .60 to .70 (Group II), .20 to .30 (Group III) and .10 to .20 (Group IV). Although these *P*-values are lower than the rest, with two exceptions they satisfy statistical norms. In other words, the differences in end terms in each age group are of less significance than the differences between the most and next most transferred relationship. This accounts for four of the five differences of questionable significance. The remaining instance occurs in Group IV for the difference between the synonym and antonym, where the *P*-value is .10 to .20.

As a further test of the existence of the gradient and of its consistency, each word, homophone, antonym and synonym was compared within each age group. The comparison was made in terms of percent of EDR transferred from the original word to its semantic relata. The tabulation of reversals, that is, EDR change in contradiction to the gradient indicated in Tables III and IV, is shown in Table V. It should be noted that the data for this table are derived from the means of the population within each age category.

From inspection of the table, it will be seen that there were only seven instances out of a total of 80 possibilities where the gradient did

TABLE V

REVERSALS OF GRADIENT IN INDIVIDUAL WORDS

Age Group	Items Compared					
	Word-homo.	Word-anto.	Word-syno.	Homo.-anto.	Homo.-syno.	Anto.-syno.
I	0	0	0	1	0	1
II	0	0	0	1	0	2
III	0	0	0	1	0	1
IV	0	0	0	0	0	0

not hold for individual stimuli. None of the reversals occurred when the original conditioned stimulus was compared to its semantically related transfer words. Furthermore, no reversals occurred where the initial and final members of the gradient were compared. It is obvious, therefore, that the trends established in the preceding paragraphs were consistent for the isolated stimuli as well as for the group data. This is additional evidence that the words selected were approximately of the same value for the four groups of *Ss*.

In summary of this part of the experiment, then, it is possible to conclude with considerable assurance that there exists a gradient of semantic relationships for stimuli in the homophone, antonym and synonym categories, and that this gradient of transference varies with the age of the subjects. For the youngest group, transfer of EDR is greatest when words sound alike and is diminished as the meanings become more similar or better as meaningful relationships enter. For the middle age groups, oppositeness and similarity tend to assume equal value. It is not until one reaches the upper teen-age groups that likeness of meaning tends to become the dominant semantic pattern.

DISCUSSION

Previous experimenters have established the possibility of studying semantic relationships by use of conditioning techniques, particularly those involving mediated generalization. Hull, Hilgard and Marquis, and others have pointed out that mediated generalization works along a gradient of stimulus intensity and magnitude. The studies of semantic conditioning have demonstrated another dimension of the mediated gradient, namely, the inherent meaningfulness of the stimuli. However, the authors of these experiments on verbal conditioning have for the most part neglected the specific study of the genetics of the gradients established by the experiments. Foley and his coworkers did show in one study (**5**) the effect of the occupation and profession of the *Ss* on the nature of the generalization patterns from neutral words to stimuli reflecting the occupational interests of the

Ss. They have also shown (**6**) that the previous educational experience of the *Ss* determined the gradient of generalization established to various lists of words. With these exceptions, however, the verbal stimuli have been implicitly held to have of themselves attributes which determined their place in the sequence of more or less favorably transferred responses. The present experiment has demonstrated that the relative strength of the semantic gradients does not depend on any a priori quality of the stimulus, but upon the way in which the whole organism utilizes language in its development. The change in the ranking of the homophones, antonyms and synonyms reflects the mode of the linguistic background of the various age groups. In other words, semantic conditioning does not depend solely on any attribute of the stimulus as such, but the stimuli must be interpreted as part of large situational and experiential wholes within which the organism exists and has learned to use and interpret verbal stimulation. This point of view has a close similarity to that advanced by Zener (**17, 18**) in his study of conditioned salivary responses in dogs operating under varied motivational conditions and freed from the restraints of the traditional holders and restricted movements of the classical Pavlovian laboratory. Here, too, the dynamic interrelationships between the bell, food, organic status of the animal and the past history of the organism contributed to an understanding of the obtained variability in response after conditioning.

The present experiment suggests also that it may be possible to study the folk-etymology emphasized by Freud and more recently by the Lifwynn group under Burrow and Galt. The latter investigator has written a stimulating paper (**7**) on the etymological implications of social neurosis. He points to etymological relations which suggest psychological conflict situations such as the correspondence between words like *double*, *doubt* and *duplicity*, *integrity* and *untouchable*, *mind*, *mental* and *mendacious*, *converse* (meaning *speak*) and *converse* (meaning *opposite*). The use of mediated generalization studies at different cultural levels and under different environmental stresses may adduce more substantial evidence than etymology for the linguistic similarities apparent in the form and structure of the words.

Summary

1. Four groups of *Ss* differing only in chronological age were conditioned to give an EDR to selected verbal stimuli.

2. Each stimulus was then followed by words representing homophonic, antonymic and synonomic relationships.

3. Transfer of EDR conditioning from the original conditioned stimuli to the semantically related words was noted.

4. For the youngest group, mean age 7–9, the homophone received most transferred EDR, the antonym and synonym following in the order stated.

5. For the age group, 10–8, the antonym stands highest, followed by the homophone and synonym.

6. For the age group, 14–0, and for the age group, 18–6, the synonym takes first place, followed by the antonym and homophone.

(Manuscript received July 18, 1945)

Bibliography

1. Cofer, C. N., & Foley, J. P., Jr. Mediated generalization and the interpretation of verbal behavior: I. Prolegomena. *Psychol. Rev.*, 1942, **49**, 513–540.
2. Cofer, C. N., Janis, M. G., & Rowell, M. M. Mediated generalization and the interpretation of verbal behavior: III, Experimental study of antonym gradients. *J. exp. Psychol.*, 1943, **32**, 266–269.
3. Diven, K. Certain determinants in the conditioning of anxiety reactions. *J. Psychol.*, 1937, **3**, 291–308.
4. Foley, J. P., Jr., & Cofer, C. N. Mediated generalization and the interpretation of verbal behavior: II, Experimental study of certain homophone and synonym gradients. *J. exp. Psychol.*, 1943, **32**, 168–175.
5. Foley, J. P., Jr., & MacMillan, Z. L. Mediated generalization and the interpretation of verbal behavior: V, 'Free association' as related to differences in professional training. *J. exp. Psychol.*, 1943, **33**, 299–310.
6. Foley, J. P., Jr., & Mathews, M. A. Mediated generalization and the interpretation of verbal behavior: IV, Experimental study of the development of inter-linguistic synonym gradients. *J. exp. Psychol.*, 1943, **33**, 188–200.
7. Galt, W. Our mother tongue. *Psychoan. Rev.*, 1943, **30**, 241–262.
8. Goodwin, J., Long, L., & Welch, L. Generalization in memory. *J. exp. Pscyhol.*, 1945, **35**, 71–75.
9. Keller, M. Mediated generalization: the generalization of a conditioned galvanic skin response established to a pictured object. *Amer. J. Psychol.*, 1943, **56**, 438–448.
10. Razran, G. H. S. Salivating and thinking in different languages. *J. Psychol.*, 1935–1936, **1**, 145–151.
11. Razran, G. H. S. Semantic, syntactic and phonetographic generalization of verbal conditioning. *Psychol. Bull.*, 1939, **36**, 578.
12. Razran, G. H.S. Transportation of relational responses and generalization of the CR. *Psychol. Rev.*, 1938, **45**, 532–538.
13. Razran, G. H. S. A quantitative study of meaning by a conditioned salivary technique (semantic conditioning). *Science*, 1939, **90**, 89–90.
14. Razran, G. H. S. Conditioned responses. *Arch. Psychol.*, 1935, No. 191.
15. Riess, B. F. Semantic conditioning involving the GSR. *J. exp. Psychol.*, 1940, **26**, 238–240.
16. Wylie, R. Generalization of semantic conditioning of the g.s.r. Unpublished M.A. thesis, Univ. Pittsburgh, 1940, p. 40.
17. Zener, K. The significance of behavior accompanying conditioned salivary secretion for theories of the conditioned response. *Amer. J. Psychol.*, 1937, **50**, 384–403.
18. Zener, K., & McCurdy, H. G. Analysis of motivational factors in conditioned behavior: I, The differential effect of changes in hunger upon conditioned, unconditioned and spontaneous salivary secretion. *J. Psychol.*, 1939, **8**, 321–350.

INCIDENTAL LEARNING AND GENERALITY OF SET

BY LEO POSTMAN AND VIRGINIA L. SENDERS *

Harvard University

Introduction

When an *S* comes to a psychological laboratory to take part in a learning experiment, he is, as a rule, explicitly instructed in what he is to do. Usually the *E* also tries to improve his *S*'s performance by providing him with some kind of incentive. *E*s have noticed, however, that *S*s would learn even if there were no instructions and no incentive. Their performance would not be so good as under an instruction to learn, to be sure, but some learning would nevertheless undeniably take place. To such learning without intent to learn the name *incidental learning* was given. McGeoch defined incidental learning as "learning which apparently takes place without a specific motive or a specified formal instruction and set to learn the specific material in question" (4, p. 299). In both 'incidental' and 'intentional' learning the *S*'s receptors must be stimulated, he must see or hear the stimulus material. The essential difference is a difference in *set*. Incidentalness refers to the absence of an experimentally induced set to learn.

An understanding of incidental learning, therefore, hinges on an understanding of the functioning of set in learning situations. Unfortunately, the concept *set* is one of the most loosely and indiscriminately used in psychology. Among the varieties of set that have been postulated at one time or another, Gibson (1, p. 781) lists: mental set, motor set, neural set, unconscious set, preparatory set, task-set, situation-set, goal-set, attitude, directing tendency, determining tendency, etc., *ad libitum*.

All these concepts, however, have certain features in common. They all imply a state of preparedness, a readiness to respond selectively to a circumscribed range of stimulus material. They all function in directing and determining associations. It is in this sense that Külpe and the searchers for the elusive mental elements of imageless thought used the term, *set*. So did Ach and later Lewin when they tried to prove that frequency and contiguity per se are not sufficient conditions of association, but that associations occur only as a result of a mental set.

* The authors wish to thank Prof. Edwin G. Boring for his advice and criticism.

Sets, however, may be 'not only determiners of associative habits but themselves habits' (1, p. 788). You learn by virtue of a mental set to learn, but you also learn to have a mental set. That learned habits may become dynamic is, e.g., implied in Woodworth's contention that mechanisms can become drives and in Allport's principle of functional autonomy.

It thus seems possible to subsume most of what has been said about set under these two generalizations:

(1) We need a concept to refer to the fact that preparedness or intention (usually experimentally induced by means of instruction to the *S*) is an important determinant of the formation of associations. That concept is *set.*

(2) Sets not only determine the course of learning but may themselves be the result of previous learning.

Such sets may have varying degrees of generality. Once a set has been experimentally induced, it will not be restricted in its operation to the specific stimuli to which the instructions refer. The set may favor (or inhibit) the learning of many related things. The generality of set to learn is illustrated by an experiment of Lepley's (3) who investigated students' memory for the names of a class roll-call as a function of serial position relative to the *S*'s own name. He found a positively accelerated gradient preceding the *S*'s own name and a decelerated gradient following his name. These data were interpreted as confirming 'in both direction and shape the demands of Hull's goal gradient hypothesis.'

Learning may be said to be *incidental* if it occurs in the absence of set. *E*s on incidental learning have tried to insure absence of set by not instructing the *S*s to learn, believing that they then had no set to learn. We can, however, make no such assumption. Sets, not induced by the *E*, may still cause the *S* to learn at least part of the stimulus material. It is plausible to believe, in view of the high degree of generalization of which sets are capable, that uncontrolled sets to learn may be operating in what appears to be incidental learning. When in an experimental situation *S*s have a general orientation toward active response, "within the general set many fleeting and specific ones may come and go" (4, p. 305).

The experimental results are clear. *S*s learn without explicit instruction to learn. Interpretations, however, differ. Is there an uncontrolled set operating in such a situation? As long as the operational definition of incidentalness remains *negative* (in terms of failure to instruct the *S*), the argument must remain inconclusive. It is necessary systematically to *control and vary* the factor of set and to find the concomitant variations in 'incidental' learning, i.e., in the learning of material other than that toward which the subject is

oriented. It is especially necessary to recognize that set is not to be approached in all-or-none terms. A learning set may be *more or less* operative with respect to a given range of stimuli, and the less effective the set the less of the stimulus material will be learned.

Perhaps the concept of *incidental learning* can be rendered unnecessary if learning be regarded as the result of (1) explicit instructions to learn a specific material. We may call the resulting set an *explicit set.* Explicit set may be induced by the experimental instructions or by self-instruction; or (2) the extension of a more general set to learn to the specific material in question. We may call such a set a *covert set.* A covert set may vary in strength, depending on the degree of transfer from the more general set. Since all learning involves activity, and hence some kind of active preparatory set, it is parsimonious to dispense with the concept *incidental learning* and to consider learning always as a function of explicit sets to learn or covert sets derived from sets of greater generality. To do so is simply to recognize that learning is always the result of an antecedent 'priming' of the organism, that it is never indeterminate or haphazard. The contention that learning is never incidental can be proved experimentally by showing that such learning is a systematic function of generality of set.

The present investigation is an experimental analysis of the generality of set. Specifically, we may ask, with the learning material and recognition tests held constant and instructions (and thereby explicit sets) varied systematically, what degree of generality does each of these experimentally induced sets have (what covert sets does each of them arouse)? How much of the stimulus material other than the material to which the instruction refers is learned? And how much of the stimulus material is learned without *any* experimentally induced set, and how does the absence of instructions compare with a graded series of systematically induced sets? The absence of explicit set, the condition under which incidental learning has been typically studied, is thus compared with a series of experimentally induced sets of varying degrees of generality.

Procedure

The *S*s were 50 undergraduates at Harvard and Radcliffe Colleges. The material used was a 350-word mimeographed selection from Tchekov's short story, *The Bet.* This material was chosen because its dramatic quality and large number of details made it especially suitable for the purposes of this experiment.

Five sets of instructions were used:

A. The *S*s were asked to read the selection for purposes of timing since the "*E* is interested in the *S*s' reading speed." As soon as they had read the selection they were instructed to raise their hands and to turn their papers over. This condition duplicates the procedure used in most experiments on incidental membory, since no explicit set to learn was established.

B. The *S*s were told: "This is a memory experiment. Read this selection carefully. You will be tested for your *general comprehension* of the material."

C. The instruction was again as in B, except that *S*s were told that they would be tested for the *specific sequence of individual events* in the story.

D. *S*s were told they would be tested for *details of content.*

E. *S*s were told they would be tested for *details of wording.*

F. *S*s were told they would be tested for *details of physical appearance.* A certain number of intentional errors, such as typing mistakes and misspellings had been introduced into the text for this purpose.

After the instructions had been read the *S*s were given a signal to start. Although every one read at his own speed, there was a time limit of 90 sec. which, however, was not exceeded in any case.

After the *S*s had finished their reading, they were occupied for three min. with a mechanical task (drawing four-in. lines) in order to eliminate rehearsal as much as possible, and were then tested by means of a multiple choice questionnaire which contained 50 questions. Of these 50 questions 10 tested general comprehension of the material (B); 10 tested memory for the sequence of events in the story (C); 10 tested for memory of details of content (D); 10 tested for memory of details of wording (E); 10 tested for memory of physical details (F). Each question was followed by four alternative responses and the *S*s were required to check the correct one. To each instruction there corresponds, therefore, a set of questions. We have two parallél continua, both ranging from the most general to the most particular: a continuum of instructions (experimentally induced sets) and a continuum of questions. When an *S* is asked to answer a question which he had been instructed to expect, we are testing an explicit set. When an *S* is asked to answer questions other than those which he had been instructed to expect, we test the effects of a covert set. The more numerous and effective the covert sets aroused by a given explicit set, the greater the generality of that explicit set.

The categories along our two continua are necessarily somewhat arbitrary. There are, e.g., no definite criteria delimiting questions testing general comprehension from questions testing sequence of events. We had, however, working criteria for delimiting our categories along a scale ranging from the most general to the most specific. A question was regarded as more general if a larger number of words in the text had to be understood and remembered in order to answer it. The criteria for categories E and F (details of wording and details of physical appearance) are fixed by their definitions. It may be that the edges of our categories have remained a little fuzzy; nevertheless we have had a rough ordinal scale of generality ranging from the most general questions to the most specific.

Results

The questionnaires were scored in terms of the number of items correctly recognized in each question category. Table I shows the

TABLE I

Percentage of Correctly Recognized Items of each Question Category under each Instruction

Question-category	Instruction					
	A	B	C	D	E	F
General comprehension	60	75	83	57	65	81
Sequence of events....	50	65	46	79	44	55
Details of content.....	49	62	54	56	37	53
Details of wording.....	34	31	33	36	30	33
Physical appearance...	31	26	26	31	38	39

percent of items correctly recognized in each category under each condition. The change in the number of correct responses in each category under each of the instructions was then determined. A direct comparison of the number of correct responses as a function of

the various instructions was impossible, since in each group different Ss, who were not matched for memory ability, were used. We took, therefore, the average number of correct responses in *all* categories for a given condition as our starting point. The number correct in each individual category was then expressed as percent deviation from the Grand Mean. Thus, if the grand average for a questionnaire was 5, and the average number of correct responses for category B was 7, this was scored as a deviation of 40 percent, etc. Fig. 1 shows percent

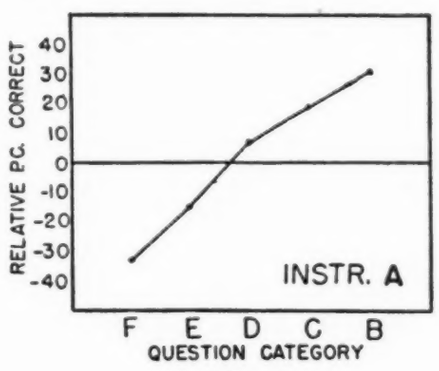

Fig. 1

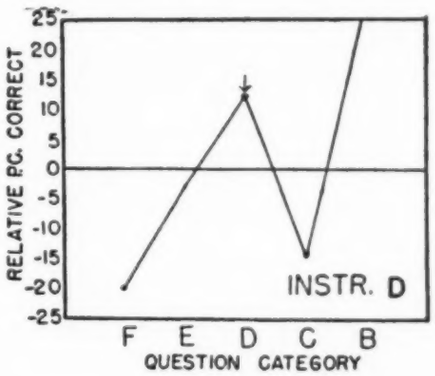

Fig. 4

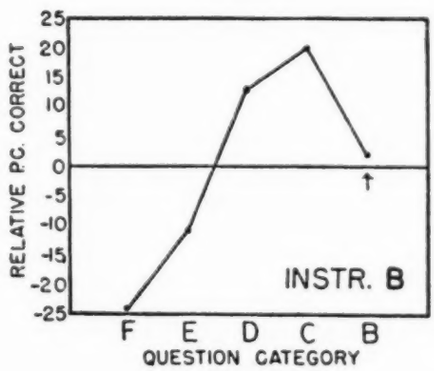

Fig. 2

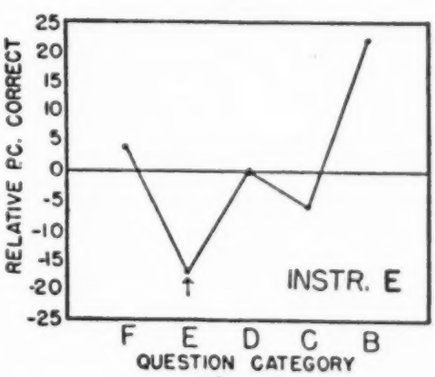

Fig. 5

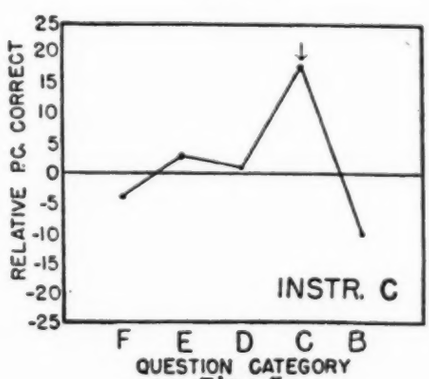

Fig. 3

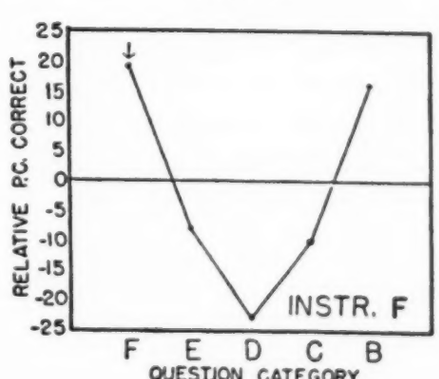

Fig. 6

Question Categories and Instructions:
Key

A No instructions
B General comprehension
C Sequence of events
D Details of content
E Details of expression
F Details of presentation

FIGS. 1–6. Change in the number of correct responses in each category under each of the instructions

deviations from the Grand Mean of each of the categories under instruction A: absence of explicit set.

Since the purpose of this experiment was to compare the various experimentally induced explicit sets with the absence of an explicit set, all other results are expressed with reference to instruction A. Under each of the instructions, we ask, does a score in a given category deviate more or less from *its* Grand Mean than does the comparable score of condition A from the Grand Mean of A? The percent deviation of each category from the Grand Mean under instruction A was, therefore, subtracted from the percent deviation of each category score from its own Grand Mean, and the results were plotted in Figs. 2–6 (categories B–F).

To illustrate with an example: Under instruction A 31 percent more items are answered correctly in question-category B, General Comprehension, than for the questionnaire as a whole. Under instructions B 34 percent more are answered in question category B. Comparing the two percent deviations, we obtain 3 percent. The ordinate in Figs. 2–6 is, therefore $(X - M) - (X_A - M_A)$, where:

X = percent of items correct in a category,
M = mean percent of items right in a questionnaire,
X_A = percent of items right in a given category under instruction A,
M_A = mean percent of items right under instruction A.

The graphs represent, therefore, the relative increase or decrease in the scores in each of the question-categories as a function of the changing instructions. They show (*a*) the effect of an explicit set (denoted by arrow) and (*b*) the effects of covert sets, i.e., sets to learn which were aroused concurrently with a given explicit set and which caused materials other than those to which the instruction referred to be learned.

This graphic presentation is purely descriptive. There is no mathematically meaningful base line, since there are no scale units. Going from left to right we have increasing generality but we cannot go beyond such *a priori ordinal* scaling. Such an account is adequate for a purely descriptive visual representation of trends.

Inspection of the graphs leaves no doubt that scores vary as a function of (*a*) different instructions, (*b*) different question-categories and (*c*) the interaction of questions and instructions. Are these variations significant? In the graphic presentation individual differences among *S*s are lost. These differences may account for at least part of the variation in the scores. By means of an analysis of variance we tested the null hypothesis that all the scores were obtained from the same parent population and the differences between the scores were due only to 'chance' factors, i.e., individual differences among *S*s

and other unknown, uncontrolled factors. The results of the analysis of variance are summarized in Table II. We may reject the null

TABLE II

Summary of Results of Analysis of Variance

Source of Variation	d.f.	Mean Variance	*F*	*P*
Instructions	5	6.472	14.947	<.01
Questions	4	13.113	30.284	<.01
Interaction	20	4.914	11.349	<.01
Error	220	.433		
	249*			

* Each of the 50 *Ss* had a score in each of the five question categories so that 250 individual scores were used in the analysis of variance, giving 249 degreees of freedom.

hypothesis at a .01 level of confidence, and conclude that the recognition scores vary systematically as a function of instruction, question-category, and interaction of instruction and question category.

Discussion of Results

The experimental results have justified our interpretation of 'incidental learning' in terms of explicit and covert sets. If 'incidental learning,' defined as learning material without an explicit instruction, were to mean that such learning occurs without any set at all, either explicit or covert, then we should expect the learning outside of a specific explicit instruction to occur uniformly or, if the data are few, at random. We find, however, that learning outside of the specific explicit instruction bears a definite demonstrable relation to the particular specificity of the explicit instruction. This relation is not necessarily one of facilitation: the important finding is that learning outside of explicit instruction is not haphazard but varies systematically with the explicit instructions. If learning of 'extraneous material' still varies systematically with the momentary explicit set and occurs from demonstrable covert sets, there is no reason to believe that learning is ever independent of set.

How do covert sets vary as a function of explicit sets? Let us consider each of the conditions in turn.

Condition A: no instruction to learn. Any learning that takes place under this condition is presumably due to the functioning of covert sets. Fig. 1 shows that, in the absence of specific instructions to learn, questions testing general comprehension of the material are answered best. On the whole, the more specific the questions, the fewer are the items correctly recognized. In the absence of specific instructions to single out any particulars for careful memorizing,

people are oriented toward general comprehension of the material before them. They need not be instructed to try to understand the things they see and hear around them. Such learning can hardly be called 'incidental.' It occurs under the learned set to order and understand the stimuli in our environment. This covert set is directed primarily toward *general* understanding and to a much lesser extent toward specific details.

Condition B: the *S*s were specifically warned that they would be tested for general comprehension. Despite this explicit set for general comprehension the number of questions correctly answered in that category *did not* increase. The explicit set for general comprehension did not prove more effective than the *S*'s covert set for general comprehension which functioned without experimental instruction. Making the set for comprehension more explicit, however, has improved memory for sequence of events. In a learning situation the *S* concentrates more on remembering the specific events of the story. The covert set for details of content is also strengthened by the learning instruction. On the other hand, details of wording and details of physical appearance are remembered less well. The *S*'s range of attention and comprehension is limited, as has been known ever since the experiments in Külpe's laboratory on attributes of sensation (**2**).

Condition C: the *S*s were instructed to concentrate on the sequence events in the story. The explicit set for sequence of events leads to a considerable improvement of retention for items in that category. Fewer questions in the category of general comprehension are answered correctly. Again this may be due to the similarity of categories B and C and the *S*'s limited range of attention and comprehension. One category is favored at the expense of the other. Neither details of content nor details of expression are favored by this instruction. The explicit set to remember sequence of events does not carry over sufficiently to these detailed features of the story to increase retention for them. There is a slight decrease in the retention for physical details of presentation. Again emphasis on content is achieved at the expense of attention to details of presentation.

Condition D: the *S*s were instructed to memorize details of content. The explicit set for details of content improves retention for items in this category. An even more striking improvement occurs in category B (general comprehension). This may be the result of a compensatory reaction on the part of the *S*. Instructed by the *E* to watch out for disjointed details, he refuses to accept this instruction without qualification. He is determined to get the general meaning of the story, to perceive it as a coherent whole. All he has time for is (**1**) to comply with the *E*'s instructions by learning details of content and (**2**) to get the 'general drift' of the story. He does so at the expense of all other

features of the story. Therefore, it is not unexpected to find that there is a decrease in the recognition scores for all other categories.

Condition E: the *S*s were instructed to concentrate on details of wording. The result of this instruction is at first surprising. Specific instructions to remember details of wording result in a considerable *decrease* in retention for these items. Why should an explicit set be an inhibitor in this case? We find an answer by referring to the so-called Von Restorff phenomenon in memory (**5**). In a series of experiments with nonsense syllables, figures, numbers, and letters, Von Restorff proved conclusively that homogeneity of material is a factor that inhibits the memory function. A 'crowding of traces' occurs if the material to be learned consists of items which are all of one kind. In our experiment, instructions to remember as many of the details of wording as possible caused a 'crowding' of a great many elements which were all similar in kind. As in the case of Von Restorff's experiments, homogeneity of material leads to a drop in retention. How does an explicit set for details of wording affect retention for the other categories? There is again a striking improvement in category B (general comprehension). Again an explicit set for small details leads to a 'compensatory' concentration on the general features of the story. The *S*'s '*Gestaltdrang*' is too great to permit him to memorize a disjointed series of words. As in the other cases, improvement in general comprehension is achieved at the expense of items in category C (sequence of events). There is, finally, an improvement in memory for physical details of presentation. An explicit set for one type of details here carries over to include other details.

Condition F: the *S*s were instructed to remember as many of the details of physical appearance of the stimulus material as they can. In contrast the the results obtained under condition E, instructions to remember details led to improved retention of these details. The type of detail, however, is radically different in these two cases. The details of physical appearance are not *similar in kind*, nor are they as numerous as the details of wording. On the contrary, the details of physical appearance are odd and unusual features which stand out 'figurally' from the rest of the stimulus material. Details of physical appearance are favored under specific instructions to remember them, but only at the expense of all other details for which retention *decreases* under condition F. We are again confronted with the phenomenon of a limited range of attention and comprehension, so that successful concentration on one type of detail can be achieved only at the expense of other details. Again we find, under condition F, the 'compensatory' effect to which we referred before. Instructed to concentrate on details of physical appearance, the *S* complies, but he is unwilling to forego his omnipresent set for understanding, his urge

to make his experience coherent and meaningful. Thus, while memizing a specific set of details, he also successfully attempts to improve his general comprehension of the material.

Reliability of differences.—In the preceding discussion the changes in retention as a function of instruction were demonstrated with reference to condition A, the zero ordinate of Figs. 2–6. We must now examine the question how reliable these changes are, i.e., which of the points on the curves in Figs. 2–6 are significantly greater than zero. To answer this question a series of *t*-tests was performed on the null hypothesis that the ordinate values in our figures are merely chance deviations from the zero ordinate. The results of these *t*-tests are summarized in Table III, which gives the values of *t* obtained as well

TABLE III

RELIABILITIES OF THE CHANGES IN RETENTION AS A FUNCTION OF INSTRUCTION

The changes in all categories are with respect to condition A, the zero ordinate of Figs. 2–6. The numbers refer to the values of *t* obtained, the numbers in parentheses give the *P*-values as obtained from Fisher's distribution of *t*.

Conditions Compared	Question Category				
	B	C	D	E	F
A–B	.59(.60)	1.83(.10)	1.91(.10)	1.53(.20)	1.10(.30)
A–C	1.48(.20)	2.25(.05)	.14(.90)	.50(.70)	.33(.80)
A–D	2.84(.02)	1.49(.20)	2.24(.05)	.84(.40)	1.37(.20)
A–E	1.98(.10)	1.75(.10)	.55(.60)	2.54(.05)	1.40(.20)
A–F	2.22(.05)	1.16(.30)	1.37(.20)	1.90(.10)	2.37(.05)

as the levels of significance at which the null hypothesis may be rejected. In view of the small number of Ss in each group, the results obtained were, of course, highly variable, but nevertheless the trends of our results, as read from Figs. 2–6 were substantiated by this statistical check. The results of the *t*-tests will now be briefly examined for each of the instructions.

Instruction B.—Retention for items in question category B did *not* increase since the *t*-test allows rejection of the null hypothesis only at the 60 percent level. The hypothesis that the improvement in categories C and D is due to chance factors may be rejected at the 10 percent level. The decrease in retention for categories E and F is less clearly reliable, the levels of significance being 20 percent and 30 percent respectively.

Instruction C.—The improvement in category C is reliable at the 5 percent level. The loss in category B is reliable at the 20 percent level. As was pointed out above, neither question category D nor question category E is favored by this instruction: the levels of confidence are 90 percent and 70 percent respectively. The slight de-

crease in cateogry F which the graph suggests does not prove statistically reliable (*P*-value 80 percent).

Instruction D.—The improvement in category D under instruction D is statistically reliable at the 5 percent level. The improvement in category B is even more clear-cut, viz., at the 2 percent level. These two changes embody the main results of instruction D. The decreases for items in the other categories are less definite as the moderately high *P*-values indicate. It should be borne in mind, however, that in working with small samples *P*-values of 20 percent are by no means negligible.

Instruction E.—The decrease in retention for items in category E is reliable at the 5 percent level, the concomitant improvement in category B is reliable at the 10 percent level, as is the decrease for items in category C. The increase for items in category F is reliable at the 20 percent level. Nothing definite may be said about question category D.

Instruction F.—The improvement in category F is reliable at the 5 percent level, the concomitant loss for category E at the 10 percent level. The decrease for category D is reliable at the 20 percent level. The 'compensatory' improvement in category B is statistically reliable at the 5 percent level. No statistically reliable change occurs for items in category C.

Conclusions

We may draw certain general conclusions from the analysis of our data:

Our results confirm the hypothesis that learning and memory are never restricted to those materials which the *S* has been instructed to learn. But the learning which occurs over and above intentional learning is not incidental in the sense in which the term *incidental* has been used in the past, viz., implying haphazardness. Incidentalness, like the concept of chance, is merely a euphemistic term for ignorance of causal conditions. For incidentalness and chance a more thorough analysis of types of set and their relative effectiveness must be substituted.

Explicit instructions in a social (e.g., laboratory) situation to learn specific material, lead to an explicit set to learn. But such a socially induced explicit set is only one of several sets to learn that may cause learning to occur.

The set to learn may be explicit but not socially induced. The *S* recognizes that certain things are worth learning or that it is to his advantage to learn them. In such a situation there is explicit self-instruction to learn. Explicit self-instructions are in turn the result of early established general sets to understand and structure environmental stimuli.

The set to learn may be covert because neither social instructions nor self-instructions are given. But covert sets are indirectly demonstrable through experimentation even though the *S* is not aware of what they are.

The results of self-instructions and covert sets have appeared to some *E*s as random or incidental because the causal conditions operating in the learning situation were partly unknown. We have reason to believe, however, that the extension of a set to learn to materials not included in *E*'s instructions varies systematically with the type of instruction and the relation in which the explicitly learned material stands to the 'extraneous' material present in the stimulus field. When there are several types of material that can be learned, specific instructions to learn a given category will improve retention for that category (with the notable exception of category E discussed above). The specific instruction will at the same time strengthen memory for some related categories, weaken others. When tests of recognition are scaled on a continuum ranging from the most general comprehension of material to memory for small details of the physical appearance of stimuli, the effects of specific instruction are usually of three kinds:

(1) There may be a gradient effect. The memory for items which are adjacent to the explicitly learned ones on the continuum of generality may be strengthened.

(2) There may be, on the other hand, mutual inhibition in the case of items which are too similar to each other.

(3) A compensatory effect may occur. Instructions to learn highly specific details cause the *S* to emphasize general comprehension in his perception and learning of the stimulus material.

There are probably other effects, but they are not likely to be any more 'incidental' or haphazard.

Further experimentation is needed in view of the importance of memory without specific instruction for the psychology of testimony. The man in the witness stand is usually called upon to testify on matters which he was not specifically instructed to observe and learn. In a sense his situation is like that of our *S*s under instruction A. Like these *S*s he is likely to remember the general trend of events and the atmosphere rather than specific details. The details he does learn will be in accordance with explicit self-instructions and covert sets which are determined by his personality in relation to the situation. (The determination of perception by interest and expectation is relevant here.) The sets that occur may only be called chance or incidental when one equates chance to ignorance of the man's personality or the situation, or the relation of the two. A witness tripped up by counsel on specific details is, therefore, not necessarily an unreliable

witness. If the witness was present at an accident, he may be quite correct in stating that driver A hit driver B and still not remember the colors of the two cars, since there was no explicit set for details of physical appearance. Reliability of a witness in respect to certain questions may make him more or *less* reliable in respect to others due to his limited range of attention and comprehension.

The laws governing the effectiveness of both explicit and covert sets are thus of importance not only for learning theory but also for the applied psychology of testimony.

(Manuscript received July 11, 1945)

Bibliography

1. Gibson, J. J. A critical review of the concept of set in contemporary experimental psychology. *Psychol. Bull.*, 1941, **38**, 781–817.
2. Külpe, O., & Bryan, W. L. Versuche über Abstraktion. *Ber. I. Kongr. exp. Psychol.*, 1904, 58–68.
3. Lepley, W. M. A gradient in incidental learning. *J. exp. Psychol.*, 1935, **18**, 195–201.
4. McGeoch, J. A. *The psychology of human learning.* New York: Longmans Green and Co., 1942.
5. Von Restorff, H. Über die Wirkung von Bereichsbildung im Spürenfeld. *Psychol. Forsch.*, **18**, 1933, 299–342.

THE LEARNING AND RETENTION OF CONCEPTS. II. THE INFLUENCE OF LENGTH OF SERIES. III. THE ORIGIN OF CONCEPTS

BY HOMER B. REED

Ft. Hays Kansas State College

The problem of this paper is to investigate (1) the influence of length of series on the learning and retention of concepts and (2) the origin of concepts. The specific problems relating to the first are to find (1) the relation of length of series to effort required to learn consistent and inconsistent concepts, (2) the relation of length of series to the percentage of consistent concepts acquired, and (3) the relation of length of series to the rate of retaining consistent concepts formed.

Method

Except for variations in quantity, the methods and materials for these investigations were the same as those reported in a previous article (5). In that investigation, a series of 42 cards was used, each of which had a nonsense syllable printed on the back and four disconnected English words on the front, one of which belonged to a group represented by the syllable. In this one, two other quantities were used, one series of 24 cards and one of 60 cards. The former consisted of the first 24 cards used in the previous investigation and the latter consisted of the entire list of 42 cards described before plus 18 new cards of the same kind. The number of concepts, six, was kept the same in the three series. Forty-two new *S*s, college students, were used, 21 in Group 6 to whom was presented the series of 24 cards, and 21 in Group 7, to whom was presented the series of 60 cards. Each group was divided into two subgroups, a and b, the members of which relearned after intervals of one week and three weeks, respectively. All groups were approximately balanced with respect to intelligence as measured by the Henmon-Nelson Test of Mental Ability for College Students. The members of each group were given Directions Number 4, which required the *S* not only to learn the names of cards but also their meaning.

II. The Influence of Length of Series

1. *The relation of length of series to the effort to learn and relearn concepts*

According to the investigations of Ebbinghaus (1), Lyon (2), Meumann (3), and Robinson and Heron (6), there is a rather consistent tendency for the rate of increase in work required to memorize increasing lengths of series of nonsense syllables to be greater than the rate of increase in the length of the series. For example, if the length of the lesson is doubled, the time required to master it is more than doubled. This is true for both meaningless and meaningful material, but the ratio between the amount of work and the amount of material is less for meaningful than for meaningless material. The curve for the results obtained by Lyon for the relation between length of rows

of nonsense syllables and the time required to memorize them in one sitting is shown in Fig. 2. The curve is parabolic in form, and contrasts sharply with the curves for learning nonsense syllables when meaning is put into them, or when they are conceptualized. When this is done, the curves in Fig. 2 and the results in Table IV show that long series are learned with relatively less effort than short ones. The number of promptings required to master a series of 60 syllables is less than twice as great as that for 24 syllables. This is true for both consistent and inconsistent concepts. But, as we saw in our previous

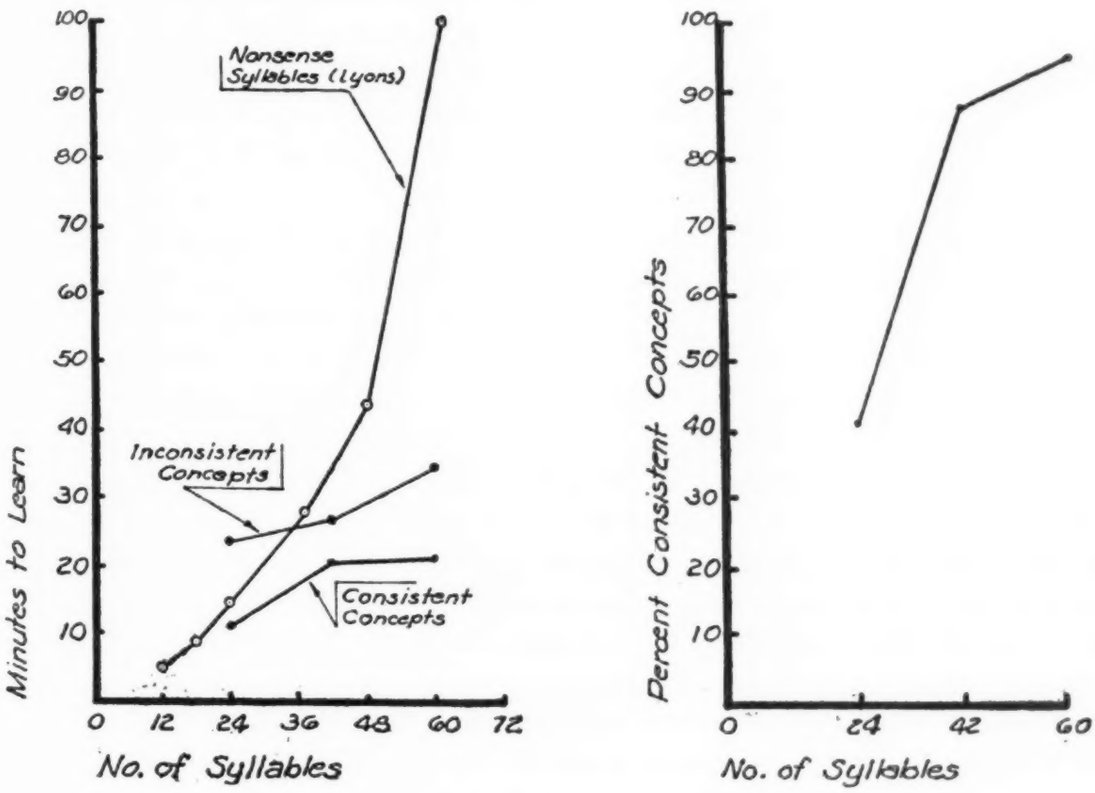

FIG. 2. Relation between length of series and amount of effort required to learn conceptualized syllables and nonsense syllables

FIG. 3. Relation between length of series and percent of consistent or correct concepts derived from them

investigation, the amount of work required to master consistent concepts is much less than that for inconsistent concepts. The length of the series also affects the rate of change in the amount of work for logical and illogical learning. This is shown by the fact that the learning curves for the two types of concepts diverge with increase in the length of the series. The curve for the inconsistent concepts is practically a straight line but bends slightly upward, while the curve for the consistent concepts bends towards the horizontal at the point for 42 syllables. In fact, the work for 60 syllables is not significantly greater than for 42 syllables.

TABLE IV

MEAN NUMBER OF PROMPTINGS PER CONCEPT TO LEARN AND RELEARN CONCEPTS IN RELATION TO LENGTH OF SERIES

No. of Syllables	Kind of Concepts	No. of Concepts	Learning		Relearning	
			Mean	S.D.	Mean	S.D.
24						
	Consistent	53	18.98	9.00	0.38	0.83
	Inconsistent	73	33.98	15.55	3.49	3.14
	Difference		15.00		3.41	
	S. E. Diff.		2.19		.51	
42						
	Consistent	98	29.19	12.05	0.91	1.61
	Inconsistent	16	40.13	17.10	2.06	2.22
	Difference		10.94		1.15	
	S. E. Diff.		4.44		.58	
60						
	Consistent	119	31.50	21.25	0.36	0.75
	Inconsistent	7	48.50	16.35	2.57	1.05
	Difference		17.00		2.21	
	S. E. Diff.		6.48		.41	

The great difference between the work curves for nonsense syllables and for concepts calls for an explanation. A defensible hypothesis for this is that learning concepts is accompanied by much less interference than memorizing mere nonsense syllables. First, the rise in the curve for nonsense syllables learned in one sitting is much sharper than for nonsense syllables learned at the rate of one reading a day, a fact which suggests that the rest periods apparently provide an opportunity for conflicting associations to be forgotten. Second, the likelihood of forming conflicting associations in the set-up of the present experiment is much less than in the memorizing of rows of nonsense syllables, in which each syllable may be associated with every other syllable in the series. In series of increasing lengths, these associations increase in geometrical order. In our set-up, the tendency is not to associate one nonsense syllable with every other one, but rather to associate the syllable with a word or a group of words on one card, the cards being distinguished from each other by the fact that each card has a group of words which is different from that on any other card. As the number of cards is increased, the number of associations for each syllable varies arithmetically rather than geometrically, providing that the *S* follows rote learning. For example, if the *S* learns the cards by associating the syllable with the first word on each card, he will have four associations for each of the six syllables in a series of 24 cards, and 10 associations for each syllable in a series of 60 cards.

We would expect the association of a syllable with 10 different words to require more work than one with only four different words; but if the words are unlike, the amount of work should not be more than $2\frac{1}{2}$ times as great. The fact that it is much less can only mean that the interference does not multiply in conceptual learning as it does with nonsense syllables alone. This accounts for the fact that the work curve for conceptual learning is nearly a straight line rather than a parabola.

Next, we have the question of deciding why the increase in amount of work in relation to the length of the series is much less for consistent concepts than for inconsistent concepts. We have already pointed out that in case of conceptual learning, one association or the concept does the work of many individual associations occurring in rote learning. Hence, the amount of work is much less. Once the concept is learned, the increase in the amount of work required to respond correctly to additional cards would not be much, for all that is required is to discriminate the key word from the others on the card and to recognize it as belonging to the concept.

The writer began this investigation with the hypothesis that the amount of work required to learn consistent concepts would decrease with increase in the length of the series. He reasoned that increasing the number of new situations in which a certain constant appeared would facilitate the isolation and definition of that constant. We shall see later that it does favor the definition of the constant, but it does not appear to reduce the amount of labor required to define it. If this investigation had been stopped when only 10 *S*s of each of Groups 6 and 7 had done the experiment, the writer's original hypothesis would have been verified but later subjects spoiled it. The majority of the early *S*s in Group 6 resorted to illogical and nonsensical methods, while those in Group 7, for the most part, used logical and relational learning. Among the later *S*s this trend was reversed. For an *S* who searches for logical relationships, a long series favors quick learning because it furnishes more opportunities for confirming or rejecting a trial hypothesis, but for one who tries to memorize the names of the cards by contiguous word, letter, and sound associations, the long series creates bewilderment and puts the subject into a hopeless muddle. When he realizes the situation, he abandons nonsensical methods and tries logical relationships, which usually bring him quick success. Such an *S* has a long learning time but comes out with consistent concepts. In contrast, an *S* who starts with logical learning also comes out with consistent concepts but has a quick learning time. Such procedures as this account for the fact that the amount of scatter as measured by the standard deviation is much greater for the logical learning for the 60-syllable series than it is for the shorter series.

2. *The relation of length of series to the percentage of consistent concepts formed*

Fig. 3 shows the relation between the percentage of consistent concepts formed and the length of the series. The percentage rises from 42 for 24 syllables through 86 for 42 syllables, to 94 for 60 syllables. This demonstrates the fact that the greater the number of new situations in which a certain constant or concept appears, the more likely it is to be observed and the more precisely will it be defined. The reasonableness of this may be seen from the *kun* cards which were as follows: No. 1 (horn, leaf, monkey, debt), No. 8 (fame, ought, tiger, saucer), No. 18 (line, people, elephant, sound), No. 20 (uncle, fried, sheep, pear), No. 30 (crowd, sail, deer, string), No. 34 (horse, circle, paid, scholar), No. 40 (carrying, died, cow, ruler).

The first three situations would favor the concept *wild animal.* The last two, the concept *domestic animal*, the last five the concept, *people*, the last four the concept, *verb* or *action;* and all of them the concept, *animal.* Since the learner is disposed to find a concept that fits all cards having the same name, he is likely to reject all possibilities except the concept, *animal.*

3. *The relation of length of series to the retention of concepts*

The results of Ebbinghaus (1), Radosawljewitsch (4), and of Robinson and Heron (6) have shown that after a constant interval the percent of work saved in relearning rows of nonsense syllables of different lengths increases with the length of the series. Is the same true for series arranged for the derivation of concepts? And is it the same for consistent and inconsistent concepts?

TABLE V

RETENTION OF CONCEPTS IN RELATION TO LENGTH OF SERIES

No. of Syllables	One Week		Three Weeks		Six Weeks	
	Consistent	Inconsistent	Consistent	Inconsistent	Consistent	Inconsistent
24	99.6	90.1	96.8	88.7		
42	97.6	95.4	95.6	91.2	96.0	89.2
60	99.5	98.1	98.3	92.0		

Table V shows that there is no constant relationship between length of series and retention of consistent concepts, but for inconsistent concepts the longer the series, the greater the amount of work saved in relearning. The relationship between length of series and retention discovered for nonsense syllables holds for inconsistent concepts only. A very important difference, however, is in the amount retained. After an interval of 24 hours, Ebbinghaus saved only 49 percent of

his readings in relearning rows of 24 syllables each while our results show a saving of 96.8 percent after three weeks for consistent concepts and a saving of 88.7 percent for inconsistent concepts. Some reasons for the difference were given in a previous paper (5).

III. Factors Influencing the Origin of Concepts

How new ideas are born is not yet fully understood, but psychologists are pretty well convinced that the process is no exception to the law of causality. Even though we cannot fully account for it, we can point to definite factors that influence the appearance of new ideas. One of these is obviously exercise, repetition or effort to learn. We have already seen that the formation of concepts under the conditions of this experiment requires a certain number of trials and promptings and that the average number of promptings varies according to conditions. One of the most important of these is what may be called the individuality of the learner. Fig. 4 shows the learning curves for four

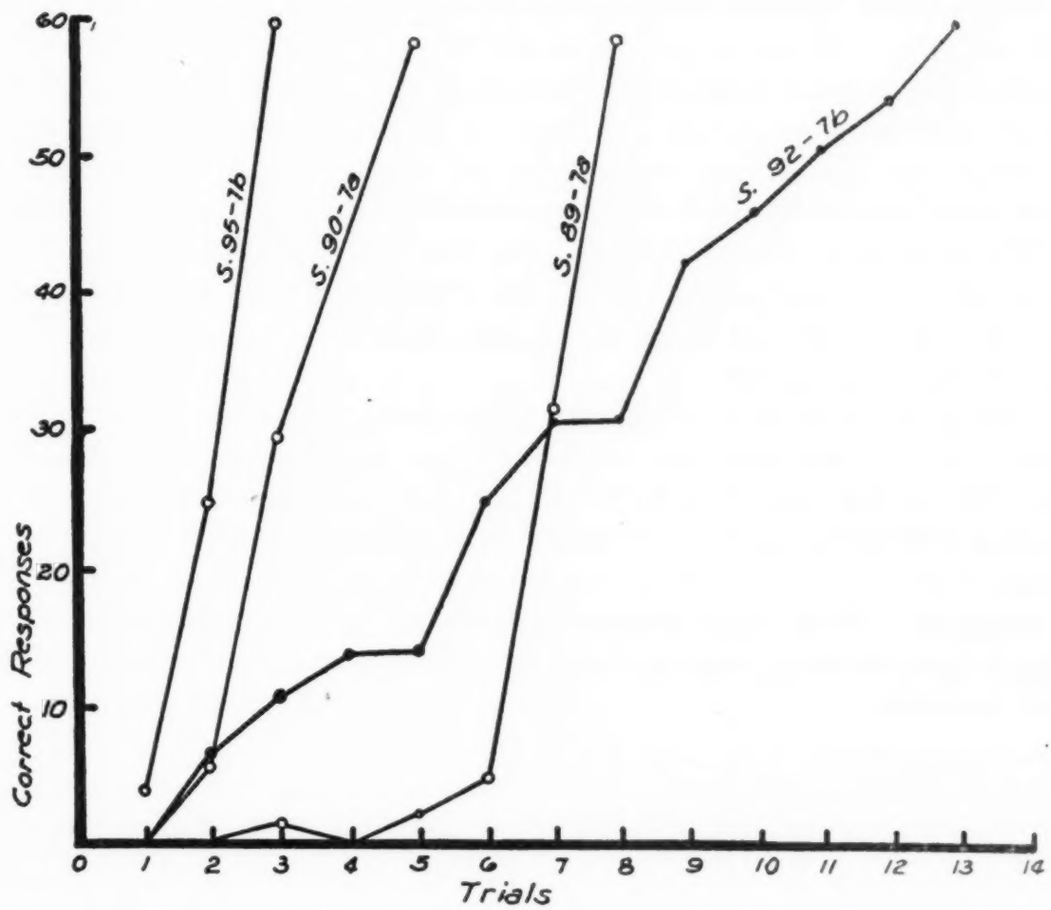

Fig. 4. Individual learning curves for responding correctly to each of 60 cards

individuals. These show a variation in trials from 3 to 13. They have a variation in total promptings from 92 to 397, and a variation in time from 24 to 120 min. The quickest of these Ss got all the

concepts correct, while the slowest one got only three correct, two partly correct, and one entirely incorrect. All required some promptings, but the promptings provided only a condition for thinking and not a guarantee for correct thinking. One factor that influenced the rate of learning was the method used. The two fast learners took their cue directly from the instructions and began at once to look for class words, while the two slow ones tried letter associations for a long time and then shifted to class terms.

Other conditions of forming concepts are generalization, knowledge of success, and the requirements of the task. An *S* may associate two or three key words belonging to a group with a given syllable, but unless he recognizes the group and generalizes that all the terms of that group belong to a syllable he does not form the concept. After he succeeds in getting the class concept for one syllable, he may or he may not generalize that the other syllables also are class terms. If he generalizes that they are, he usually has rapid success. If he does not, he has learned nothing from success with one syllable that helps him with another. Knowledge of success was provided by the *E*'s saying, 'Right,' when the *S* named the card correctly, and his prompting with the correct syllables if the *S* failed. This was the principal criterion by which the *S* affirmed or rejected his hypothesis. The acceptance of the task provided the *S* with the necessary motivation for his efforts.

We shall now discuss the factors that are particularly related to the origin of consistent concepts and then those that are particularly related to the origin of inconsistent concepts.

1. *Consistent concepts*

One factor that favored the generalization that the syllables represented class terms was the *set* created by the instructions. Instructions No. 4 tell the *S* definitely that the task is an experiment in learning concepts, that a concept is a name for any one of a group of things, that each nonsense syllable is a concept, and that he is to find its meaning. With such definite instructions, we would expect the *S* to start immediately looking for class terms, but this is true only of a small percent.

For example, Subject 77–6b seems to have fully comprehended the instructions. Immediately after hearing them, she said, "Do you mean that there is a certain group of words that goes with each syllable?" She completed the task in an average of 13.5 promptings against an average of 27.70 for her group, and had all the concepts correct. Subject 95–7b, whose curve is shown in Fig. 4, evidently started with the same idea, for she began to respond correctly during the first trial and had all the concepts except one correct at the end of the second trial. Subject 90–6a, whose curve is also shown in Fig. 4, began with class terms. She remarked: "First I got a word, then I put it in a class, and the next time I would see if the class was right."

The *frequency* of the words belonging to a given class was a factor that favored the suggestion of the class concept. The following observations illustrate this factor:

Subject 88–7a. "I noticed *horse* was *kun*, also *dog* and *cat*, and then thought *kun* was animal. I noticed too that the *vor* cards had such words as *sweetheart*, *girlie*, *sweetie*, etc. and so thought *vor* was love."

Subject 65–6a. "I memorized the first card, *kun*, and associated it with *monkey*, and then I saw that the other *kuns* had animals on th cards. I saw the names of flowers were on the *yem* cards. I first associated *dax* with words beginning with *b* as in *blue*, but *b* didn't work with other cards. Then I tried colors."

Subject 75–6b. "First I looked for a word in a certain position but soon gave it up. Then I tried fitting one of the words with each syllable. Then I noticed that color words went with *dax*. Then I looked for other kinds of words for each of the other syllables."

Identical elements between syllables and words sometimes led to the consistent concept.

Subject 20–5a. "First I noticed *love*, turned *ov* around and that made it *vo* as in *vor*. *Vor* is *love*. The first time through I noticed some flowers, some trees, some colors, and some animals. I supposed there might be a syllable for each, and watched out for the connections."

Subject 43–4a "*Kun* is a German word for *can*. Animals can do things, so *kun* is an animal. My roomate goes with a girl named Lavora, so I associated *vor* with love."

Subject 93–7b. "My first procedure was to associate syllables and letters. I missed too often but caught *kun* and *cat*, *kun* and *cow*, then I saw *dog*, *elephant*, and *lion* on *kun* cards, and finally decided that *kun* is animal."

Sometimes *contiguity* between a syllable and a key word gave the successful lead.

Subject 21–5a. "I got on it through orange. I learned that *orange* was *dax* and then I saw that other *daxes* were colors."

If the *S* failed to get all the consistent concepts during the learning period, he often grasped them quickly during the relearning period. Questioning brought out that the correct concepts were discovered by *reflection* during the rest interval.

Subject 74–6b. "I learned last time that *bep* was *vegetables*. During the interval, I decided that the other syllables represented groups of things. Today, I kept a watchout for the groups and learned what they were."

More frequently not a single factor but a *combination* of them, such as identical elements, frequency, contiguity, and passive observation, led to correct concepts.

Subject 89–7a. (See curve in Fig. 4) Trial 3. No correct concepts. "I was completely puzzled and wondered if you were repeating the same syllables over and over. I tried to associate all the words on one card with the syllables. I also tried to associate the letters of a certain word with the syllable, for example; *kun* and *cow*. I tried to associate the syllable with word position; for example, first, second, third, and fourth."

Trial 5. "*Kun* is *cow*, *vor* is *noisy*."

Trial 6. "I tried not to think of anything, leave myself open and see if anything would come, but not much came. *Kun* makes me think of *o's*. *Vor* makes me think of *ors*. This is supposed to give me concepts?"

Trial 8. All concepts correct. "In two or three *kuns*—I saw a *cow* on one, a *dog* on another, and *deer* on another. Then I decided that *kun* was animal. About the same time, I discovered that *yem* was flowers and *dax* color. Finally, I discovered that *jik* was tree."

2. *Inconsistent concepts*

To find the factors that influence the origin of inconsistent concepts, we tabulated the inconsistent or incorrect concepts for 23 *S*s from each of Groups 4 and 5 given in response to the *E*'s questions: "What suggests [name of syllable]? What is [name of syllable]? And in what ways have you tried to learn these syllables? An additional purpose in using the records from these groups was to discover the influence of set or of instructions on the number and kind of errors.

After tabulating the errors for each syllable, they were classified into groups which in the judgment of the experimenter fitted them and which brought out as far as possible the psychological factors influencing their origin. After several trials, the classification shown in Table VI seemed to be most satisfactory.

TABLE VI

CLASSIFICATION OF INCONSISTENT CONCEPTS IN RELATION TO SET IN 23 SUBJECTS FROM EACH OF GROUPS 4 AND 5

Group 4 was set to learn names and meanings, Group 5 was set to learn names only.

Type of Responses	Percent Group 4	Group 5
Response from proper card		
Similar to proper response or stimulus (e.g. *beet* for *bep*, *cow* for *kun*)	12.40	17.22
Response from wrong card		
Similar to proper response or stimulus (e.g. *crawl* for *kun*, *picnic* for *jik*)	1.53	3.33
Not similar to proper response or stimulus (e.g. *brook* for *dax*)	6.58	2.37
Response not in list		
Similar to proper response or stimulus (e.g. *cigarette* for *cigar*)	3.54	10.18
Not similar to proper response or stimulus (e.g. *far* for *yem*)	4.56	.78
Wrong category in list (e.g. *flowers* for *dax*)	3.80	4.50
Too broad a category (e.g. *food* for *bep*)	8.35	5.48
Position in proper card		
First word	28.86	36.40
Second word	8.35	2.35
Third word	6.58	4.70
Fourth word	11.65	6.65
Position in series (e.g. *Kun* is the first card)	.51	4.31
Miscellaneous	3.29	1.76
Total Per Cent	100.00	100.00
Total Number	395	511

Studying the table from the standpoint of determiners for inconsistent concepts, we find that about 40 percent are related to position habits. The concept is either the first, second, third, or fourth word; and of these over half of the errors is determined by the first word on the card. The psychological factor here may be called the *primacy of a contiguous experience.* That is, the first word that an *S* sees in connection with a new name is the most important single determiner of an erroneous meaning for the name. The second most frequent source of error is *similarity in sensory elements* either between stimulus and response or between correct response and wrong response. From

17 to 30 percent of the erroneous concepts may be attributed to this factor. Some word on the card that has a common letter or sound with the response syllable or with one of the stimulus words is associated with the response syllable or its meaning. The third most frequent error is *misgrouping*. From 10 to 12 percent of the errors are due to this factor. The new name is thought of as representing a group of things, but the group may be too broad or may be put in one of the groups occurring in the list but the wrong one, or it may be placed in too narrow a group. For example, *bep* instead of representing vegetables is thought of as standing for food, or it may be thought of as standing for colors which is the proper class for *dax*, and *dax* may be thought of representing *fruits* for which there is one word on the *dax* cards and one on the *yem* cards. The fourth most frequent group of errors, which comprises from 8 to 11 percent, is classified as not similar to proper response or stimulus. These consist of words which have no apparent correlation with either response or stimulus but seem to be floating around in the *S*'s mind and are given because of that fact. They may be thought of as due to *perseveration*.

Erroneous concepts are also related to the *S*'s set. This is obvious from the fact that the errors in Group 5, set to learn names only, are much more numerous than in Group 4, set to learn names and meanings, 511 as against 395. Lack of specific set or instructions is not only a fertile source of errors, but also favors certain kinds of errors, namely, those related to primacy, serial position, and similarity in the form of common letters or sounds such as a *y* word for *yem* or a *b* word for *bep* or *dax*. It is easy to see how such concepts would lead to wrong responses, for *y* occurs not only in *yem* words like *daisy* and *lilly* but also in a *kun* word like *monkey* and in *jik* words like *sorry* and *eyes*. Particular letters or sounds do not occur consistently with any one syllable. The importance of primacy is seen in the fact that the percentage of error for the first word on the card is one-third higher for Group 5 than for Group 4; and of serial position, in the fact that the percentage of error relating to it is four times as high in Group 5 as in Group 4. But, creating a specific set through instructions to get both the name of each card and its meaning also has a fly in the ointment. The percentage for the error called *too broad a category*, is nearly twice as high for Group 4 as for Group 5. This is understandable in view of the fact that a member of Group 4 is disposed to look for groupings. Often the group found and chosen is too broad, negative instances being ignored.

Another factor that determines inconsistent concepts is the *recency* of the stimulus. In the responses to the *E*'s questions, "What is *kun?*" or "What is *jik?*" he observed that many *S*s would name the first word from the last *kun* or *jik* card seen; second, a word from next

to the last *kun* or *jik* card seen, and so on until the last word given was taken from the first card shown. That is, there appeared to be a tendency to name the words for which a given syllable stood in the order of the recency with which they were experienced. An effort was made to measure this factor. Possibilities were found in (1) a correlation between the *recency position* of each wrong response and the *recency order* of the presentation, (2) a correlation between the *recency position* of the response of the first word on a card and the *recency order* of the presentation, and (3) a correlation between the *frequency* of the first response of the first word on a card and the *recency order* of presentation. One objection to the first measure is that we have primacy or the position factor interfering with the recency factor. Some *S*s have a habit of looking first at the right end or the middle of a card rather than the left end. In such cases there are at least two variables at work. The second and third measures have the advantage of keeping the position factor constant, and were therefore used for trial computations for the tabulations of all the wrong responses or inconsistent concepts given by 23 *S*s from Group 4 and the same number from Group 5. Both the measures used showed positive correlations with the recency factor. In Group 5, the number of first responses from the first word of the card showed 17 for the seventh or most recent card and 4 for the first or least recent card, the rank correlation between the number of first responses from the first word and the recency of presentation being .63. The corresponding correlation for Group 4 was .89. In Group 5, the average recency position from the first words on the card was 1.57 for the seventh or most recent card, and 2.62 for the first or least recent card, the rank correlation between the two factors being .50. In Group 4, the corresponding correlation was .86. These results show that the recency of the experience of a syllable and word is a factor in producing a wrong concept of the syllable. We would not expect the correlation to be perfect, because some *S*s make an effort to learn in the order or presentation and are also influenced in recall by such factors as familiarity, interest, word length, and similarity elements, none of which were strictly controlled.

Comparing the factors that lead to the origin of consistent and of inconsistent concepts, it appears that they are the same for both types—namely set, identical elements, frequency, contiguity, and generalization. The difference lies in the kind of process represented by these terms. Consistent concepts are favored by a set directed towards group classification, focalized generalization, identical elements based on meaning or group relationships, and contiguity and frequency in so far as they relate to a given logical group. Inconsistent concepts are favored by a general rather than a specific set,

undirected generalization, identical perceptual elements, the primacy of a contiguous word, and its recency.

Implications

The facts reported in this paper have certain implications of a practical nature with reference to length of task assigned and the improvement of the accuracy of thinking. The facts on the relation between the amount of work and the length of the material suggest that if the task is such as to require the mastering of a certain number of meaningful facts which have for the learner little systematic relation to each other, the amount of work increases proportionately with the amount of material; but the ratio of increase in the work is less than the ratio of increase of the material. In other words, a double lesson means something less than a double amount of work. If the material is meaningless, then the ratio of increase in the work is greater than the ratio of increase in the material, and a double lesson means more than a double amount of work. But if the material may be logically organized by the learner, and if there is a certain constant number of problems to be solved from a choice of varying lengths of materials, then the amount of work increases with the amount of material only up to a certain point, after which an increase in the amount of material means only an insignificant increase in the amount of work. As regards improving the accuracy of thinking, the most important implication is that thinking may be improved. The facts of this experiment suggest that the accuracy of a concept may be increased by (1) giving the *S* instructions that will create a specific set in favor of the concept, (2) keeping the goal constantly in mind, (3) persisting until a solution is obtained, (4) increasing the number of examples that illustrate the concept, (5) modifying a hypothesis when contrary cases are met, (6) questioning the first or most recent, or a frequent concomitant factor, and (7) questioning a concomitant factor which has a sensory similarity.

Summary

In this paper two general problems are reported upon: (1) the influence of length of series on the learning and retention of concepts and (2) the origin of concepts. The materials consisted of three series of cards, one of 24, one of 42, and one of 60. One of six nonsense syllables appeared irregularly on the back of each card while on the face were printed four English words, one of which belonged to a logical group represented by the syllable. The *S*'s task was to discover this logical group and name each card correctly as the cards were presented at the rate of one every seven sec. For the solution of the first prob-

lem, a total of 72 *S*s were used, and for the solution of the second problem, a total of 99 subjects. The principal conclusions derived from the results are the following:

1. Within the limits of this experiment, the amount of work required to master a long series is relatively less but absolutely greater than for a short series. For inconsistent concepts there is a straight line relationship between the amount of work and the length of series. For consistent concepts this is true only up to a certain point beyond which an increase in length of series does not appear to increase significantly the amount of work.

2. The percentage of consistent concepts formed increases directly with the number of different situations in which concepts appear.

3. There appears to be no consistent relationship between the retention of correct concepts and the length of series.

4. The retention of incorrect concepts increases consistently with the length of the series.

5. General factors that influence the origin of concepts are effort to learn, generalization, knowledge of success, and motivation or the requirements of the task.

6. The amount of effort required varies immensely from one individual to another.

7. Specific factors that influence the formation of consistent concepts are set, frequency of key words, tendency to generalize, identical elements between syllables and concept words, contiguity between a syllable and key word, reflection after learning, and a combination of several of these sometimes with passive but more often with active observation.

8. Specific factors that influence the formation of inconsistent concepts are about the same, but important determiners of illogical concepts are primacy of contiguous factors or position habits, recency of contiguous factors based on identical sensory elements between syllables and content words, failure to generalize, and perseveration.

The difference in the determiners of consistent and inconsistent concepts lies in the form of each. Consistency concepts are favored by a set directed toward group classification, focalized generalization, identical elements based on meaning or group relationships, and contiguity and frequency as they relate to a given logical group. Inconsistent concepts are favored by a general rather than a specific set, undirected generalization, identical perceptual elements, primacy of contiguous factors, and their recency.

(Manuscript received July 31, 1945)

References

1. Ebbinghaus, H. *Memory: a contribution to experimental psychology.* (Trans. by H. A. Ruger and C. E. Busenius.) New York: Teachers College, Columbia Univ., 1913.
2. Lyon, D. O. The relation of length of material to time taken for learning and the optimum distribution of time. *J. educ. Psychol.*, 1914, 5, 1–9; 85–91, 155–163.
3. Meumann, E. *Psychology of learning.* (Trans. by John Wallace Baird.) New York: D. Appleton and Co., 1913.
4. Radosawljewitsch, P. R. *Das Behalten und Vergessen bei Kindern und Erwachsenen nach experimentellen Untersuchugen.* Leipzig: Nemnich, 1907.
5. Reed, H. B. The learning and retention of concepts. I. The influence of set. *J. exp. Psychol.*, 1946, 35, 71–88.
6. Robinson, E. S., & Heron, W. T. Results of variations in length of memorized material. *J. exp. Psychol.*, 1922, 5, 428–448.

DISCUSSION

COMMENTS ON CRITERIA OF EASE OF READING

BY MATTHEW LUCKIESH

Director Lighting Research Laboratory, General Electric Company, Nela Park, Cleveland

In a recent issue of this JOURNAL (1), M. E. Bitterman released some work under the title of "Heart Rate and Frequency of Blinking as Indices of Visual Efficiency." In his summary he states that his results do not support the conclusions of Luckiesh and Moss. We have found that many factors influence the frequency of blinking. Inasmuch as the author does not describe the test conditions adequately, it is quite possible, and even quite probable, that factors which he has not taken into account may have resulted in conclusions different from ours.

Dr. Bitterman's experiment with type-size involves physical work as well as reading. So does McNally's experiment which requires the underlining of a word after reading a short statement and question (2). Furthermore, McNally's *Ss* were children with subnormal vision. We used educated adults with normal vision who were used to being subjects. This is a major difference which probably leads to a major difference in the results.

The effect on the blink-rate of discontinuous reading punctuated by manual operations cannot be directly compared with the results of our work which was always done with as little interruption of the reading as possible. The difficult nature of the tasks in Dr. Bitterman's, as well as McNally's experiments, required strict concentration which we might well expect to inhibit the act of blinking. In our work we aimed in every way to have conditions such that the subjects read in a *natural manner under normal conditions*. We consider this of utmost importance. As for the heart-rate, the performance of the physical work would likely control that and overshadow any effect of the visual effort. Furthermore, we have made no claims as to the effect of type-size on the heart-rate.

Although McNally used 72 school children and was careful to arrange the tests so that differences resulting from order of presentation were minimized, the experiment was performed only once for each *S*. In our experiments we have found that the five-min. experiment-time for the blink-rate is often sufficient to reveal significant differences, but not when only one set of data is obtained for each *S*. At least two sets of data are always taken, and when differences are not large, several sets are often necessary for statistical accuracy.

Another major difference between our method and Bitterman's is the use of electrodes on the *S* near his eyes when counting the blink-rate. We carefully avoided anything of that sort which might affect a person's blinking although he might not be particularly conscious of it. In our work the

blinks were observed and counted by the *E*, who was not visible to the *S* while reading. We have considered this procedure to be very important.

In the discussion of his level-of-illumination experiment Dr. Bitterman does not state what the brightness of the surroundings was in each test with 3 and 91 footcandles, respectively, on the reading matter. We have found that the brightness of the surroundings influences the blink-rate very markedly. This might readily mask any effect of level of illumination. In other words, the brightness conditions in the entire visual field combine to influence the rate of blinking. The net results of various brightness factors are not readily predictable.

In regard to the heart-rate, it appears that the author has read into our use of it something which was not there. We have not claimed or tried to show that "visual effort causes cardiac deceleration." Our results indicate that when an *S* is relaxed and reading easy-to-read material in his own natural way, the heart-rate is depressed with the lower levels of illumination which we used. Dr. Bitterman's experiment confirms this. The performance of physical work is likely to control the heart-rate or, at least, to overshadow any effect of visual effort. Furthermore, we have made no claims as to the effect of type-size on the heart-rate.

Our researches with the rate of involuntary blinking covered a continuous period of about ten years with many *S*s and involved the study of *many* variables. Any work which aims to verify our results must naturally involve the same conditions, which means the same control over the various factors which influence the rate of involuntary blinking. We do not know what some of Dr. Bitterman's conditions were, and we particularly emphasize the importance of the lighting conditions, the brightness distribution in the visual field, and the brightness of the immediate surroundings of the reading matter. Anyone who has followed our work closely is familiar with the fact that all of these factors influence the rate of involuntary blinking. If one or more of them is allowed to intrude itself prominently, it is obvious that the effects could mask the effects of type-size, level of illumination or any other factor being studied.

In the course of several decades of researches in seeing we have searched for, and have extensively studied, various possible criteria of ease of seeing. Our results with the blink-rate and heart-rate are consistent with our results with other effects such as nervous muscular tension and fatigue of the extrinsic eye-muscles. Furthermore, they are consistent with the results of our extensive studies of visibility which can scarcely be controversial. Admittedly the blink-rate as a criterion of ease of reading is the result of a complexity of factors which can lead to different conclusions unless the same attention has been given to the control of various factors. We make no claims beyond those which appear justifiable in our original papers, viewed individually and as a comprehensive whole based upon continuous researches in seeing over a long period. A convenient list of references to our original papers is to be found in our recent book *Reading as a visual task* by Matthew Luckiesh and Frank K. Moss, D. Van Nostrand Co., 1942. In that volume will also be found the results of our work on speed of reading which lead to our conclusion as to its inadequacy as a criterion of ease of reading. In general,

we rely primarily on our work in visibility, but it is desirable to have a dependable criterion of ease of seeing. In our laboratory the rate of involuntary blinking has been quite promising. However, it is obvious that it must be used with an understanding of its limitations and with careful technique.

REFERENCE

1. BITTERMAN, M. E. Heart rate and frequency of blinking as indices of visual efficiency. *J. exp. Psychol.*, 1945, 35, 279–292.
2. MCNALLY, H. J. The readability of certain type sizes and forms in sight-saving classes. *Teach. Coll. Contrib. Ed.*, 1943, No. 883.

A REPLY TO DR. LUCKIESH

BY M. E. BITTERMAN

Cornell University

Dr. Luckiesh's comments on my recent attempt to evaluate some of his well-known techniques for the measurement of visual efficiency would appear to call for further discussion. It is my opinion that his criticisms are completely lacking in justification.

(1) Dr. Luckiesh contends that I have not adequately described the test conditions employed in my experiments. The only "inadequacy" of which he makes specific mention is my failure to describe the "brightness of the immediate surroundings of the reading matter," and he here implies my ignorance of his purported finding that frequency of blinking is influenced by the brightness-ratio of fixated to peripheral areas. My awareness of Luckiesh's data is indicated by a specific reference to brightness-ratio in the paper under discussion (1, p. 282). The important point, however, is that since brightness-ratio was not a variable under investigation it was necessary, not to measure it, but only to control it. In Experiment I the conditions of illumination were constant throughout. In Experiment II only the intensity of the light-source was varied, and the brightness-ratio therefore remained unchanged. If Luckiesh is contending that the blink index is valid only with certain brightness-ratios (and he has not thus far intimated that such is the case), he only succeeds in casting further doubt upon its practical value.

(2) The visual task employed in Experiment I is unsatisfactory to Luckiesh on a number of grounds—it is "unnatural," it requires periodic manual operations (marking 'plus' and 'minus' signs with a pencil), and it demands too much "concentration." His own *S*s, Luckiesh explains, "read in a *natural manner under normal conditions*," although he probably would be hard-pressed to convey the precise meaning of his italicized phrase. But in experiments dealing with visual efficiency it is necessary to have some objective check upon the performance of the *S*s (4, 5), and the most realistic experiment imaginable is of dubious value if uncontrolled in this respect.

Tinker (6) has pointed out that Luckiesh himself has often drawn conclusions based upon speed tests, although he at times categorically denounces such tests. The contention that the limited motor responses required of the *Ss* were sufficient to overshadow any effect of visual effort upon the heart-rate only serves to weaken the position of those who assert that this index is a useful one for the measurement of visual efficiency. Luckiesh's claim that the task required "strict concentration which we may well expect to inhibit blinking" is fully met by the results of a recent experiment (2) which indicates no relation between frequency of blinking and effort expenditure in mental work. Furthermore, it should be noted that even in Experiment II, in which the conditions were equivalent to those usually employed by Luckiesh and Moss, frequency of blinking did not increase with the duration of reading, nor did it vary with illumination in the manner which they describe.

(3) Luckiesh protests that he has "not claimed or tried to show that visual effort causes cardiac deceleration." I could, however, quote at least a dozen passages from his own writings to show that he does indeed postulate a causal relationship between visual effort and cardiac activity. In the very same paragraph which contains his protest, apparently referring to my failure to find decrements in Experiment I, he writes: "The performance of physical work is likely to control the heart-rate or, at least, to overshadow any effect of visual effort." In *The science of seeing* we find the following significant passage:

> Since the heart responds to numerous physiological and psychological stimuli, it seemed possible that the expenditure of human energy in critical seeing might also be revealed or indicated by changes in heart-rate. Thus if critical visual effort produced 'reverberations' in the central nervous system, which spread excitation to the peripheral system, it follows that heart-rate should be affected (3, p. 214).

Luckiesh's results are purported to demonstrate that the heart-rate is lower under conditions of illumination calling for greater expenditure of effort. While he has made no specific statement with regard to type-size, the general terms in which his postulate is phrased appear to warrant the extension of experimental work to variables other than lighting.

(4) Luckiesh's final criticism deals with the electromyographic technique of recording eyelid movements which was employed in my investigations. He believes that the active electrode placed in the vicinity of one eye (on the supra-orbital ridge) may influence the frequency of blinking. Actually, however, the effect of the electrode is negligible, as demonstrated by preliminary validation procedures. The values obtained through the use of the electromyographic method are of the same order as those recorded in short periods of direct observation. The observation method is extremely fatiguing for *E* and probably less reliable over long periods. Furthermore, my interest was in relative rather than absolute frequencies, and for this reason it was necessary only to ensure that whatever effect the electrode may have had was constant throughout each experiment.

The weight of present evidence, my own as well as that of other workers (1), is heavily against many of the findings which Dr. Luckiesh has widely publicized and defended. It is clearly *his* responsibility to reëxamine his

experimental procedures and reëvaluate the conditions under which his work was done.

REFERENCES

1. BITTERMAN, M. E. Heart rate and frequency of blinking as indices of visual efficiency. *J. exp. Psychol.*, 1945, 35, 279–292.
2. BITTERMAN, M. E., & SOLOWAY, E. The relation between frequency of blinking and effort expended in mental work. *J. exp. Psychol.*, 1946, 36, 134–136.
3. LUCKIESH, M., & MOSS, F. K. *The science of seeing.* New York: Van Nostrand, 1937.
4. PATERSON, D. G., & TINKER, M. A. *How to make type readable.* New York: Harper, 1940.
5. PYKE, R. L. *The legibility of print.* London: His Majesty's Stationery Office, 1926.
6. TINKER, M. A. Review of *Light, vision and seeing* by M. Luckiesh. *J. appl. Psychol.*, 1945, 29, 252–253.

REYNOLDS' REPETITION OF BLODGETT'S EXPERIMENT ON LATENT LEARNING

BY HUGH C. BLODGETT

University of Texas

In a recent article in this JOURNAL, Reynolds (5) states in his discussion of the results of Experiment 1, ". . . *in the present experiment no latent* learning was observed. What was observed was that the latent learning group *learned during the non-reward period.*" And, with a slight variation in wording, he repeats this conclusion in connection with his Experiment 2.

It seems to me that it would have been equally accurate for him to have said that the considerable amount of learning shown by the non-reward groups precludes any demonstration of latent learning. Learning is for all groups largely manifest.

First of all, I think it should be pointed out that a demonstration of latent learning does not necessarily rest upon any day to day comparison of groups. Broadly, the concept generalizes the fact that the non-reward group, following the introduction of reward, closes the gap both in time and error scores between itself and the control group. The non-reward curve increases in acceleration and catches up with the control curve.[1] This phenomenon has been reported many times, first by Szmanski (8), then by Simmons (7), and since by others (3, 6, 8, 9), some of them not associated with Tolman in experimental work, or in theoretical viewpoint. The effect can, of course, best be shown when a relatively difficult maze such as the Tolman-Honzik maze is used. This probably accounts for the fact that the statistical significance of my finding was less than that reported by Tolman and Honzik, and by others using a 14-unit T maze rather than the simpler six-unit Blodgett maze.

[1] The change from reward to non-reward should be made while learning is still taking place in both control and non-reward-groups, so that the 'catching-up' can occur before the control has reached terminal improvement. If the control has completed its learning before the non-reward group has been shifted to reward, comparison between groups can be made only in terms of daily rate of change in improvement, as in the Blodgett experiment.

The reality of latent learning does not depend upon Blodgett's criterion cited by Reynolds in his footnote 1, nor upon Blodgett's second criterion, which Reynolds does not mention. This second criterion was used in anticipation of the criticism which Reynolds directs against all latent learning experiments of the so-called Blodgett type. In a word, reduction in error score for the non-reward group was compared with reduction in error score *at an equivalent ordinate* for the control. This procedure would, I thought, make for a fair comparison of non-reward and control performance and would forestall such criticism as that of Reynolds' statement, "Greater decrease in errors for latent learning animals than for control animals can be taken as indication of latent learning only on the assumption that the errors are equivalent in both cases."

The assumptions underlying our second mode of comparison seem to have been brought in question by Reynolds' observations on depth of cul entry in his Experiment 2. He purports to show that the type of cul entry is characteristically different for reward and non-reward running. More work is certainly needed on this aspect, particularly in view of the fact that Reynolds control learning curve is somewhat atypical. It shows little, if any, negative acceleration, differing in this from his control group 1, and from the reward curves of other multiple-T mazes (see 4, Fig. 50). We hardly seem justified in assuming that Reynolds has given us a definite statement of the relationship between part and complete cul entry during the progress of reward learning.

In connection with Reynolds' work on degree of cul entry, a second point of some interest relates to the curves of the three-day non-reward and the seven-day non-reward groups of Experiment 2. It will be noticed in Reynolds' Fig. 4, and Table II, that no apparently significant difference in *total* errors appears at any state of learning as between these groups. This is true for the first three days when both groups are running under non-reward conditions, and for the subsequent four days when the three-day group is running under reward, and the seven-day group, non-reward. However, by Reynolds' reasoning, this apparent similarity is deceptive, since the ratio of part to complete errors is different for the two groups. The three-day group, running from days 4 through 7, under reward conditions, makes *more complete cul entries* than the seven-day group running under parallel conditions, but with no reward. The process of eliminating all but the partial errors (many of them border-line in type) seems to have been carried out more effectively by the seven-day group under non-reward than by the three-day group with reward.

This brings us back to the real discrepancy between my findings and those of Reynolds. Antecedent to any discussion of latent learning is the question—do non-reward rats show marked degrees of learning and effective performance, or both? At this point I should like to refer to a related study, that of E. E. Anderson on *externalization* of drive (1).

Anderson found, it will be recalled, that the maze learning of non-reward animals was greatly influenced by their previous training. When compared as to rate of learning on a six-unit multiple-T maze,[2] non-reward animals

[2] Anderson's maze was identical in true path sequence with the Blodgett maze.

that had previously received reward training on another maze learned almost as rapidly as a rewarded control group; whereas a second non-reward group with no previous training was greatly inferior to the 'maze wise' non-reward group. In fact, Reynolds' non-reward curves resemble Anderson's 'maze wise' or 'externalized' non-reward (see Anderson's Fig. 1); mine on the other hand, resemble Anderson's non-reward control with no previous training (see Anderson's Fig. 3). My seven-day non-reward curve shows slightly more error reduction than Anderson's during the first seven days of learning. However, his scoring was probably stricter, five inches of cul entry as against my criterion of body length.

In view of Anderson's findings, it may seem surprising that, in repeating my work on latent learning, Reynolds should have included animals that had received previous training. It is true that Reynolds' non-reward 'naive' rats performed about as well as his 'maze wise' rats. Yet this fact presents us with a further ambiguity, Reynolds' failure to observe externalization of drive as reported by Anderson. It is certainly to be hoped that further work, embodying Reynolds' useful suggestions regarding depth of cul entry, will be done, and will add to our evidence on this whole matter of non-reward learning.

Perhaps it is typical of rats in non-reward situations that they show efficient performance. However, the wealth of contrary evidence, and particularly that of Elliot and Tolman and Honzik on performance following removal of reward, makes this seem very unlikely.

(Manuscript received August 6, 1945)

Bibliography

1. Anderson, E. E. The externalization of drive: III. Maze learning by non-rewarded and by satiated rats. *J. gen. Psychol.*, 1941, **59**, 397–426.
2. Elliot, M. H. The effect of change of reward on the maze performance of rats. *Calif. Publ. Psychol.*, 1928, **4**, 19–30.
3. Herb, Frances H. Latent learning—non-reward followed by food in blinds. *J. comp. Psychol.*, 1939, **29**, 247–256.
4. Moss, F. A. *Comparative psychology.* N. Y.: Prentice-Hall, 1942. Rev. ed.
5. Reynolds, J. B. The effect of the introduction of reward upon the maze performance of rats—a repetition of the Blodgett experiment. *J. exp. Psychol.*, 1945, **35**, 504–516.
6. Spence, K. W., & Lippitt, R. O. 'Latent' learning of a simple maze problem with relevant needs satisfied. *Psychol. Bull.*, 1940, **37**, 429.
7. Simmons, R. The relative effectiveness of certain incentives in animal learning. *Psychol. Monogr.*, 1924, **2**, 1–79.
8. Szymanski, J. S. Versuche über die Wirkung der Factoren, die als Antrieb zum Erlernen einer Handlung dienen können. *Pflüg. Arch. ges. Physiol.*, 1918, **171**, 374–385.
9. Tolman, E. C., & Honik, C. H. Introduction and removal of reward, and maze performance in rats. *Univ. Calif. Publ. Psychol.*, 1930, **4**, 257–275.
10. Williams, K. A. The reward value of a conditioned stimulus. *Univ. Calif. Publ. Psychol.*, 1927, **4**, 31–55.